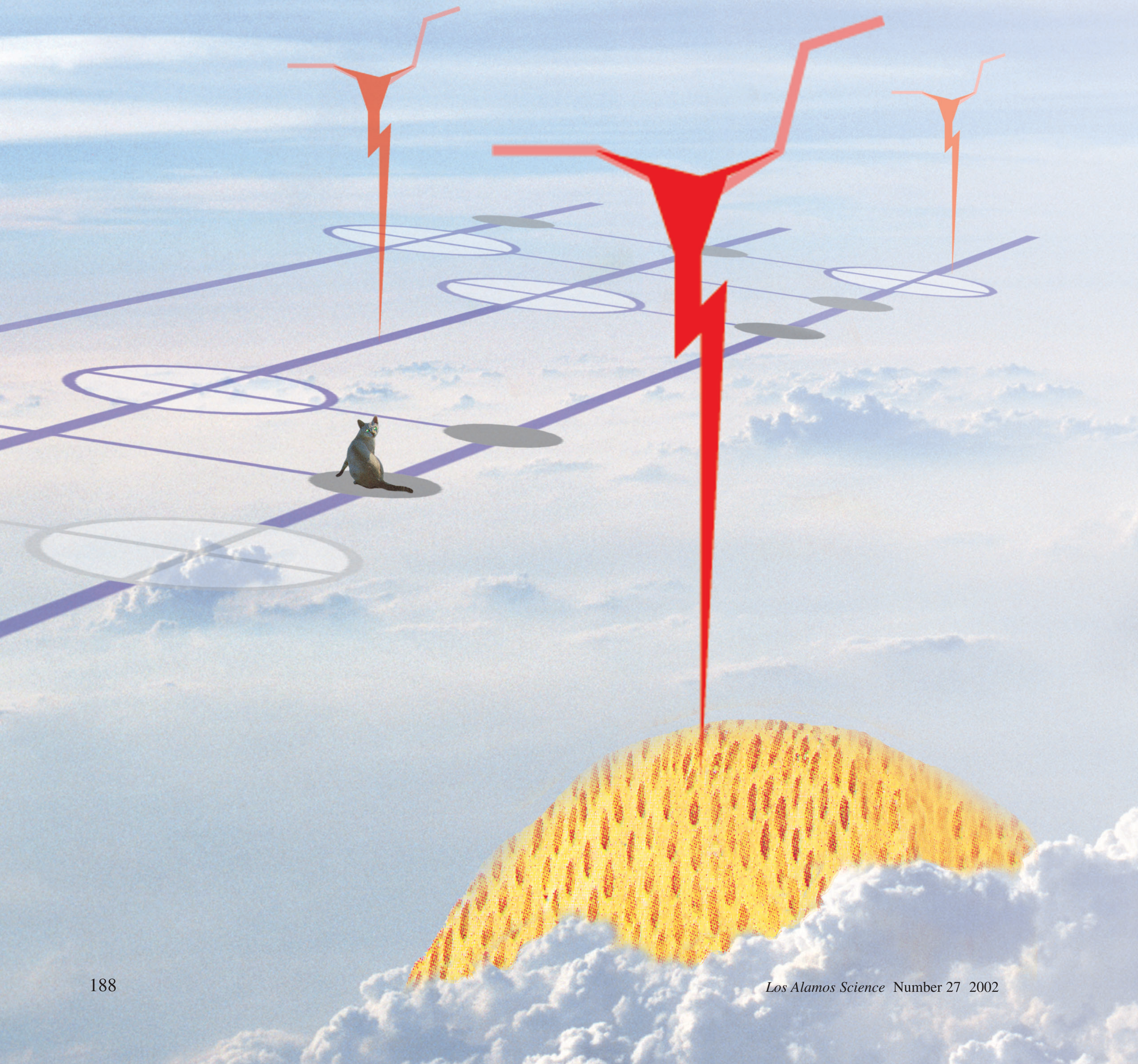
Introduction to Quantum Error Correction

Emanuel Knill, Raymond Laflamme, Alexei Ashikhmin, Howard N. Barnum, Lorenza Viola, and Wojciech H. Zurek



When physically realized, quantum information processing (QIP) can be used to solve problems in physics simulation, cryptanalysis, and secure communication for which there are no known efficient solutions based on classical information processing. Numerous proposals exist for building the devices required for QIP by using systems that exhibit quantum properties. Examples include nuclear spins in molecules, electron spins or charge in quantum dots, collective states of superconductors, and photons (Braunstein and Lo 2000). In all these cases, there are well-established physical models that, under ideal conditions, allow for exact realizations of quantum information and its manipulation. However, real physical systems never behave exactly like the ideal models. The main problems are environmental noise, which is due to incomplete isolation of the system from the rest of the world, and control errors, which are caused by calibration errors and random fluctuations in control parameters. Attempts to reduce the effects of these errors are confronted by the conflicting needs of being able to control and reliably measure the quantum systems. These needs require strong interactions with control devices and systems that are sufficiently well isolated to maintain coherence, the subtle relationship between the phases in a quantum superposition. The fact that quantum effects rarely persist on macroscopic scales suggests that meeting these needs requires considerable outside intervention.

Soon after Peter Shor published the efficient quantum factoring algorithm with its applications to breaking commonly used public-key cryptosystems, Andrew Steane (1996) and Shor (1995) gave the first constructions of quantum error-correcting codes. These codes make it possible to store quantum information so that one can reverse the effects of the most likely errors. By demonstrating that quantum information can exist in protected parts of the state space, they showed that, in principle, it is possible to protect against environmental noise when storing or transmitting information. Stimulated by these results and in order to solve errors happening during computation with quantum information, researchers initiated a series of investigations to determine whether it was possible to quantum-compute in a fault-tolerant manner. The outcome of these investigations was positive and culminated in what are now known as accuracy threshold theorems (Gottesman 1996, Calderbank et al. 1997, Calderbank et al. 1998, Shor 1996, Kitaev 1997, Knill and Laflamme 1996, Aharonov and Ben-Or 1996, Aharonov and Ben-Or 1999, Knill et al. 1998a, Knill et al. 1998b, Gottesman 1998, Preskill 1998). According to these theorems, if the effects of all errors are sufficiently small per quantum bit (qubit) and computation step, then it is possible to process quantum information arbitrarily accurately with reasonable resource overheads. The requirement on errors is quantified by a maximum tolerable error rate called the threshold. The threshold value depends strongly on the details of the assumed error model. All threshold theorems require that errors at different times and locations be independent and that the basic computational operations can be applied in parallel. Although the proven thresholds are well out of the range of today’s devices, there are signs that, in practice, fault-tolerant quantum computation may be realizable.

In retrospect, advances in quantum error correction and fault-tolerant computation were made possible by the realization that accurate computation does not require the state of the physical devices supporting the computation to be perfect. In classical information processing, this observation is so obvious that it is often forgotten: No two letters “e” on a written page are physically identical, and the number of electrons used to store a bit in the computer’s memory varies substantially. Nevertheless, we have no difficulty in accurately identifying the desired letter or state. A crucial conceptual difficulty with quantum information is that, by its very nature, it cannot be identified by being “looked” at. As a result, the sense in which quantum information can be

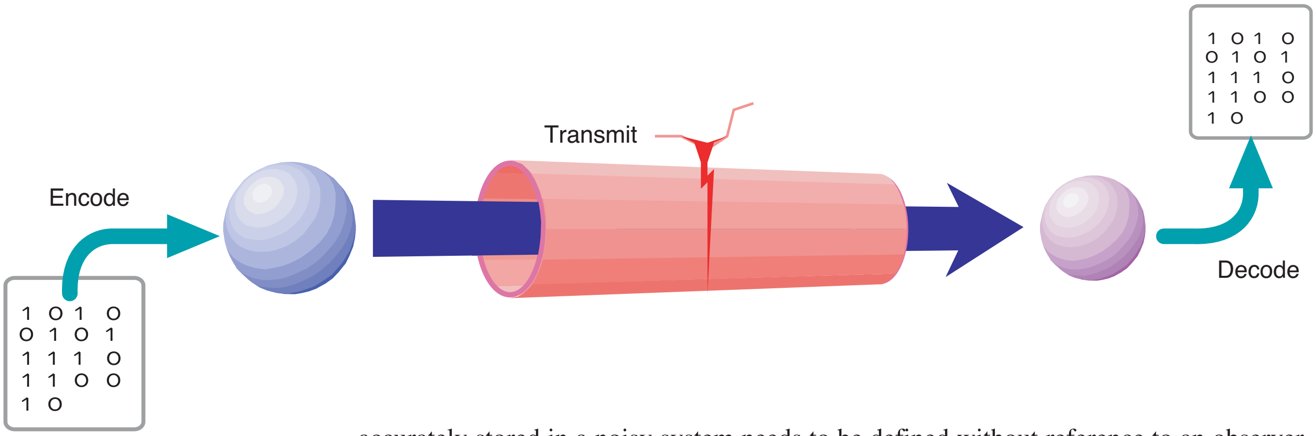


Figure 1. Typical Application of Error Correction Methods
The three main steps required for communication are shown in this figure: Information is first encoded in a physical system, then transmitted over the noisy communication channel, and finally decoded. The combination of encoding and decoding is chosen so that errors have no effect on the transmitted information.

accurately stored in a noisy system needs to be defined without reference to an observer. There are two ways to accomplish this task. The first is to define stored information to be the information that can, in principle, be extracted by a quantum decoding procedure. The second is to explicitly define "subsystems" (particle-like aspects of the quantum device) that contain the desired information. The first approach is a natural generalization of the usual interpretations of classical error-correction methods, whereas the second is motivated by a way of characterizing quantum particles.

In this article, we motivate and explain the decoding and subsystems view of quantum error correction. We explain how quantum noise in QIP can be described and classified and summarize the requirements that need to be satisfied for fault tolerance. Considering the capabilities of currently available quantum technology, the requirements appear daunting. But the idea of subsystems shows that these requirements can be met in many different, and often unexpected, ways.

Our article is structured as follows: The basic concepts are introduced by example, first for classical and then for quantum codes. We then show how the concepts are defined in general. Following a discussion of error models and analysis, we state and explain the necessary and sufficient conditions for detectability of errors and correctability of error sets. That section is followed by a brief introduction to two of the most important methods for constructing error-correcting codes and subsystems. For a basic overview, it suffices to read the beginnings of these more-technical sections. The principles of fault-tolerant quantum computation are outlined in the last section.

Concepts and Examples

Communication is the prototypical application of error correction methods. To communicate, a sender needs to convey information to a receiver over a noisy communication channel. Such a channel can be thought of as a means of transmitting an information-carrying physical system from one place to another. During transmission, the physical system is subject to disturbances that can affect the information carried. To use a communication channel, the sender needs to encode the information to be transmitted in the physical system. After transmission, the receiver decodes the information. The procedure is shown in Figure 1.

Protecting stored information is another important application of error correction methods. In this case, the user encodes the information in a storage system and retrieves it later. Provided that there is no communication from the receiver to the sender, any error correction method applicable to communication is also applicable to storage and vice versa. In a later section ("Fault-Tolerant Quantum Communication and Computation" on page 217), we discuss the problem of fault-tolerant computation, which requires enhancing error correction methods in order to enable applying operations to encoded information without losing protection against errors.

To illustrate the different features of error correction methods, we consider three examples. We begin by describing them for classical information, but in each case, there is a quantum analogue that will be introduced later.

Trivial Two-Bit Example. Consider a physical system consisting of two bits with state space {00, 01, 10, 11}. We use the convention that state symbols for physical systems subject to errors are in gray. States changed by errors are shown in red.[1] In this example, the system is subject to errors that flip (apply the **not** operator to) the first bit with probability .5. We wish to safely store one bit of information. To this end, we store the information in the second physical bit because this bit is unaffected by the errors (see Figure 2).

As suggested by the usage examples in Figure 1, one can encode one bit of information in the physical system by the map that takes 0 → 00 and 1 → 01. This means that the states 0 and 1 of an ideal bit are represented by the states 00 and 01 of the noisy physical system, respectively.

To decode the information, one can extract the second bit by the following map:

$$\begin{array}{lcl} 00 & \rightarrow & 0 \\ 10 & \rightarrow & 0 \\ 01 & \rightarrow & 1 \\ 11 & \rightarrow & 1 \end{array} \qquad (1)$$

This procedure ensures that the encoded bit is recovered by the decoding regardless of the error. There are other combinations of encoding and decoding that work. For example, in the encoding, we could swap the meaning of 0 and 1 by using the map 0 → 01 and 1 → 00. The new decoding procedure adds a bit flip to the one shown above. The only difference between this combination of encoding/decoding and the previous one lies in the way in which the information is represented in the range of the encoding. This range consists of the two states 00 and 01 and is called the code. The states in the code are called code words.

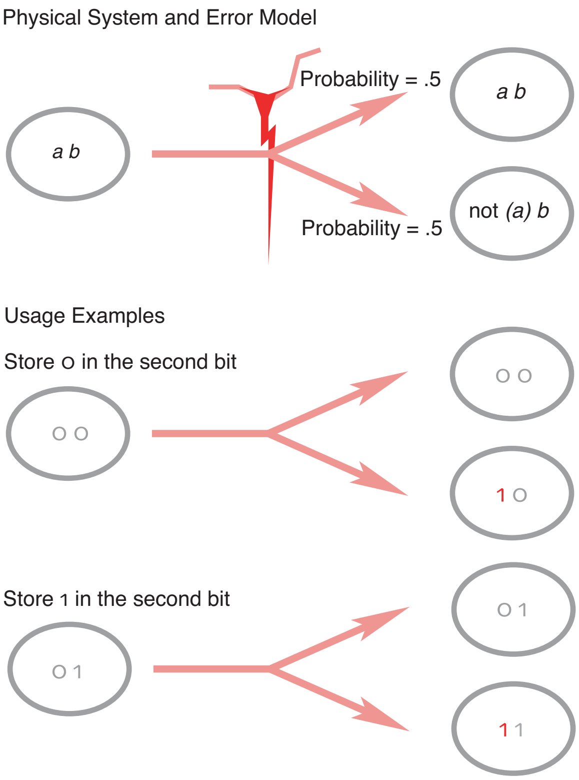

Physical System and Error Model
a b
Probability = .5
a b
Probability = .5
not (a) b
Usage Examples
Store 0 in the second bit
0 0
0 0
1 0
Store 1 in the second bit
0 1
0 1
1 1

Figure 2. A Simple Error Model

Errors affect only the first bit of a physical two-bit system. All joint states of the two bits are affected by errors. For example, the joint state 00 is changed by the error to 10. Nevertheless, the value of the information represented in the second physical bit is unchanged.

Although trivial, the example just given is typical of ways for dealing with errors. That is, there is always a way of viewing the physical system as a pair of abstract systems: The first member of the pair experiences the errors, and the second carries the information to be protected. The two abstract systems are called subsystems of the physical system and are usually not identifiable with any of the system's physical components. The first is the syndrome subsystem, and the second is the information-carrying subsystem. Encoding consists of initializing the first system and storing the information in the second. Decoding is accomplished by extraction of the second system. In the example, the two subsystems are readily identified as the two physical bits that make up the physical system. The first is the syndrome subsystem and is initialized to 0 by the encoding. The second carries the encoded information.

---

[1] These graphical conventions are not crucial for understanding what the symbols mean and are intended for emphasis only.

The Repetition Code. The next example is a special case of the main problem of classical error correction and occurs in typical communication settings and in computer memories. Let the physical system consist of three bits. The effect of the errors is to independently flip each bit with probability p, which we take to be $p = .25$. The repetition code results from triplicating the information to be protected. An encoding is given by the map $0 \rightarrow 000$, $1 \rightarrow 111$. The repetition code is the set $\{000, 111\}$, which is the range of the encoding. The information can be decoded with majority logic: If two out of three bits are 0, the output is 0; otherwise, the output is 1.

How well does this encoding/decoding combination work for protecting one bit of information against the errors? The decoding fails to extract the bit of information correctly if two or three of the bits were flipped by the error. We can calculate the probability of incorrect decoding as follows: The probability of a given pair of bits having flipped is $.25^2 * .75$. There are three different pairs. The probability of three bits having flipped is $.25^3$. Thus, the probability of error in the encoded bit is $3 \cdot .25^2 * .75 + .25^3 = 0.15625$. This is an improvement over .25, which is the probability that the information represented in one of the three physical bits is corrupted by error.

To see that one can interpret this example by viewing the physical system as a pair of subsystems, it suffices to identify the physical system's states with the states of a suitable pair. The following shows such a subsystem identification:

$$\begin{array}{lcl}
000 & \leftrightarrow & 00 \cdot 0 \\
001 & \leftrightarrow & 11 \cdot 0 \\
010 & \leftrightarrow & 01 \cdot 0 \\
100 & \leftrightarrow & 10 \cdot 0 \\
011 & \leftrightarrow & 10 \cdot 1 \\
101 & \leftrightarrow & 01 \cdot 1 \\
110 & \leftrightarrow & 11 \cdot 1 \\
111 & \leftrightarrow & 00 \cdot 1
\end{array} \tag{2}$$

The left side consists of the 8 states of the physical system, which are the possible states for the three physical bits making up the system. The right side shows the corresponding states for the subsystem pair. The syndrome subsystem is a two-bit subsystem, whose states are shown first. The syndrome subsystem's states are called syndromes. After the "·" symbol are the states of the information-carrying one-bit subsystem.

In the subsystem identification above, the repetition code consists of the two states for which the syndrome is 00. That is, the code states 000 and 111 correspond to the states $00 \cdot 0$ and $00 \cdot 1$ of the subsystem pair. For a state in this code, single-bit flips do not change the information-carrying bit, only the syndrome. For example, a bit flip of the second bit changes 000 to 010, which is identified with $01 \cdot 0$. The syndrome has changed from 00 to 01. Similarly, this error changes 111 to $101 \leftrightarrow 01 \cdot 1$. The following diagram shows these effects :

$$\begin{tikzcd}
000 \arrow[d] & \leftrightarrow & 00 \cdot 0 & 111 \arrow[d] & \leftrightarrow & 00 \cdot 1 \\
010 & \leftrightarrow & 01 \cdot 0 & 101 & \leftrightarrow & 01 \cdot 1
\end{tikzcd} \tag{3}$$

Note that the syndrome change is the same. In general, with this subsystem identification, we can infer from the syndrome which single bit was flipped on an encoded state.

Errors usually act cumulatively over time. For the repetition code, this is a problem in the sense that it takes only a few actions of the above error model for the two- and three-bit errors to overwhelm the encoded information. One way to delay the loss of information is to decode and reencode sufficiently often. Instead of explicitly decoding and reencoding, the subsystem identification can be used directly for the same effect, namely, that of resetting the syndrome subsystem's state to 00. For example, if the state is 10 · 1, it needs to be reset to 00 · 1. Therefore, using the subsystem identification, resetting requires changing the state 011 to 111. It can be checked that, in every case, what is required is to set all bits of the physical system to the majority of the bits. After the syndrome subsystem has been reset, the information is again protected against the next one-bit error.

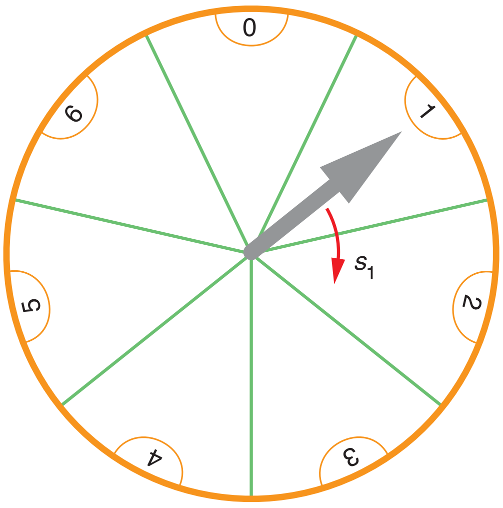


Figure 3. A Seven-State Cyclic System
The position of the pointer on the seven-position dial determines the state of the system. With the pointer in the position shown, the state is 1. Errors have the effect of rotating the pointer clockwise or counterclockwise. The effect of s_1 is to rotate the pointer clockwise, as shown by the red arrow.

A Code for a Cyclic System. We next consider a physical system that does not consist of bits. This system has seven states symbolized by 0, 1, 2, 3, 4, 5, and 6. Let s_1 be the right-circular shift operator defined by $s_1(l) = l+1$ for $0 \leq l \leq 5$ and $s_1(6) = 0$. Define $s_0 = \mathbb{1}$ (the identity operator),

$$s_k = \underbrace{s_1 \dots s_1}_{k \text{ times}}, \tag{4}$$

and $s_{-k} = s_k^{-1}$ (left-circular shift by k). The model can be visualized as a pointer on a dial with seven positions, as shown in Figure 3. Suppose that the errors consist of applying s_k with probability qe^{-k^2}, where $q = 0.5641$ is chosen so that the probabilities sum to 1, that is $\sum_{k=-\infty}^{\infty} qe^{-k^2} = 1$. Thus, s_0 has probability 0.5641, and each of s_{-1} and s_1 has probability 0.2075. These are the main errors that we need to protect against. Continuous versions of this error model in the context of communication channels are known as Gaussian channels.

One bit can be encoded in this physical system by the map 0 → 1, 1 → 4. To decode with protection against s_0, s_{-1}, and s_1, use the mapping

$$\begin{array}{lcl} 0 & \rightarrow & 0 \\ 1 & \rightarrow & 0 \\ 2 & \rightarrow & 0 \\ 3 & \rightarrow & 1 \\ 4 & \rightarrow & 1 \\ 5 & \rightarrow & 1 \\ 6 & \rightarrow & \text{fail} \end{array} \tag{5}$$

If state 6 is encountered, we know that an error involving a shift of at least 2 (left or right) occurred, but there is no reasonable way of decoding it to the state of a bit. This means that the error is detected, but we cannot correct it. Error detection can be used by the receiver to ask for information to be sent again. The probability of correctly decoding with this code is at least 0.9792, which is the probability that the error caused a shift of at most 1.

As before, a pair of syndrome and information-carrying subsystems can be identified as being used by the encoding and decoding procedures. It suffices to correctly identify the syndrome states, which we name -1, 0, and 1, because they indicate which of the likeliest shifts happened. The resulting subsystem identification is

$$\begin{array}{ccr} 0 & \leftrightarrow & -1 \cdot 0 \\ 1 & \leftrightarrow & 0 \cdot 0 \\ 2 & \leftrightarrow & 1 \cdot 0 \\ 3 & \leftrightarrow & -1 \cdot 1 \\ 4 & \leftrightarrow & 0 \cdot 1 \\ 5 & \leftrightarrow & 1 \cdot 1 \end{array} \tag{6}$$

A new feature of this subsystem identification is that it is incomplete: Only a subset of the state space is identified. In this case, the complement can be used for error detection.

Like the repetition code, this code can be used in a setting where the errors happen repeatedly. Again, it suffices to reset the syndrome subsystem, in this case to 0, to keep the encoded information protected. After the syndrome subsystem has been reset, a subsequent s_1 or s_{-1} error affects only the syndrome.

Principles of Error Correction

When considering the problem of limiting the effects of errors in information processing, the first task is to establish the properties of the physical systems that are available for representing and computing with information. Thus, it is necessary to learn the following: the physical system to be used, in particular the structure of its state space; the available means for controlling this system; the type of information to be processed; and the nature of the errors, that is, the error model. With this information, the approaches used to correct errors in the three examples provided in the previous section involve the following:

1. Determine a code, which is a subspace of the physical system, that can represent the information to be processed.
2. (a) Identify a decoding procedure that can restore the information represented in the code after any one of the most likely errors occurred or (b) determine a pair of syndrome and information-carrying subsystems such that the code corresponds to a "base" state of the syndrome subsystem and the primary errors act only on the syndrome.
3. Analyze the error behavior of the code and subsystem.

The tasks of determining a code and identifying decoding procedures or subsystems are closely related. As a result, the following questions are at the foundation of the theory of error correction: What properties must a code satisfy so that it can be used to protect well against a given error model? How does one obtain the decoding or subsystem identification that achieves this protection? In many cases, the answers can be based on choosing a fixed set of error operators that represents well the most likely errors and then determining whether these errors can be protected against without any loss of information. Once an error set is fixed, determining whether it is correctable can be cast in terms of the idea of detectable errors. This idea works equally well for both classical and quantum information. We introduce it using classical information concepts.

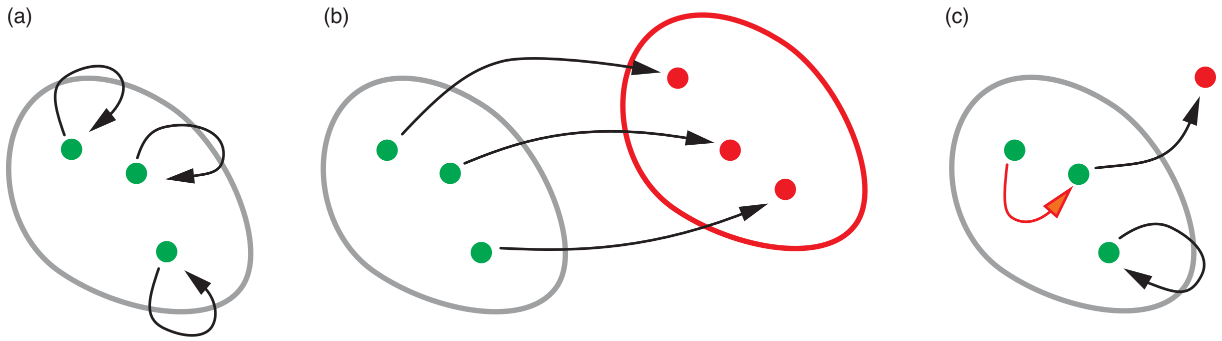


Figure 4. Typical Detectable and Undetectable Code Errors
Three examples are shown. In each, the code is represented by a brown oval containing three code words (green points). The effect of the error operator is shown as arrows. (a) The error does not change the code words and is therefore considered detectable. (b) The error maps the code words outside the code so that it is detected. (c) One code word is mapped to another, as shown by the red arrow. Finding that a received word is still in the code does not guarantee that it was the originally encoded word. The error is therefore not detectable.

Error Detection. Error detection was used in the cyclic-system example to reject a state that could not be properly decoded. In the communication setting, error control methods based on error detection alone work as follows: The encoded information is transmitted. The receiver checks whether the state is still in the code, that is, whether it could have been obtained by encoding. If not, the result is rejected. The sender can be informed of the failure so that the information can be retransmitted. Given a set of error operators that need to be protected against, the scheme is successful if, for each error operator, either the information is unchanged or the error is detected. Thus, we can say that an operator E is detectable by a code if, for each state x in the code, either $Ex = x$ or Ex is not in the code (see Figure 4).

What errors are detectable by the codes in the examples? The code in the first example consists of 00 and 01. Every operator that affects only the first bit is therefore detectable. In particular, all the operators in the error model are detectable. In the second example, the code consists of the states 000 and 111. The identity operator has no effect and is therefore detectable. Any flips of exactly one or two bits are detectable because the states in the code are changed to states outside the code. The error that flips all bits is not detectable because it preserves the code but changes the states in the code. With the code for the cyclic system, shifts by –2, –1, 0, 1, and 2 are detectable but not shifts by 3.

To conclude the section, we state a characterization of detectability, which has a natural generalization to the case of quantum information.

Theorem 1. E is detectable by a code if and only if for all $x \neq y$ in the code, $Ex \neq y$.

From Error Detection to Error Correction. Given a code C and a set of error operators $\mathcal{E} = \{\mathbb{1} = E_0, E_1, E_2 \ldots\}$, is it possible to determine whether a decoding procedure or subsystem exists such that $\mathcal{E}$ is correctable (by C), that is, such that the errors in $\mathcal{E}$ do not affect the encoded information? As explained below, the answer is yes, and the solution is to check the condition in the following theorem:

Theorem 2. $\mathcal{E}$ is correctable by C if and only if, for all $x \neq y$ in the code and all i and j, it is true that $E_i x \neq E_j y$.

Observe that the notion of correctability depends on all the errors in the set under consideration and, unlike detectability, cannot be applied to individual errors.

To see that the condition for correctability in Theorem 2 is necessary, suppose that for some $x \neq y$ in the code and some i and j, we have $z = E_i x = E_j y$. If the state z is obtained after an unknown error in $\mathcal{E}$, then it is not possible to determine whether the original code word was x or y because we cannot tell whether E_i or E_j occurred.

To see that the condition for correctability in Theorem 2 is sufficient, we assume it and construct a decoding method $z \rightarrow \mathrm{dec}(z)$. Suppose that after an unknown error occurred, the state z is obtained. There can be one and only one x in the code for which some $E_{i(z)} \in \mathcal{E}$ satisfies the condition that $E_{i(z)}x = z$. Thus, x must be the original code word, and we can decode z by defining $x = \mathrm{dec}(z)$. Note that it is possible for two errors to have the same effect on some code words. A subsystem identification for this decoding is given by $z \leftrightarrow i(z) \cdot \mathrm{dec}(z)$, where the syndrome subsystem's state space consists of error operator indices $i(z)$ and the information-carrying system's consists of the code words $\mathrm{dec}(z)$ returned by the decoding. The subsystem identification thus constructed is not necessarily onto the state space of the subsystem pair. That is, for different code words x, the set of $i(z)$ such that $\mathrm{dec}(z) = x$ can vary and need not be all the error indices. As we will show, the subsystem identification is onto the state space of the subsystem pair in the case of quantum information. It is instructive to check that, when applied to the examples, this subsystem construction does give a version of the subsystem identifications provided earlier.

It is possible to relate the condition for correctability of an error set to detectability. For simplicity, assume that each E_i is invertible. (This assumption is satisfied by our examples but not by error operators such as "reset bit one to 0.") In this case, the correctability condition is equivalent to the statement that all products $E_j^{-1} E_i$ are detectable. To see the equivalence, first suppose that some $E_j^{-1} E_i$ is not detectable. Then, there are $x \neq y$ in the code such that $E_j^{-1} E_i x = y$. Consequently, $E_i x = E_j y$, and the error set is not correctable. This argument can be reversed to complete the proof of equivalence.

If the assumption that the errors are invertible does not hold, the relationship between detectability and correctability becomes more complicated, requiring a generalization of the inverse operation. This generalization is simpler in the quantum setting.

Quantum Error Correction

The principles of error correction outlined before apply to the quantum setting as readily as to the classical setting. The main difference is that the physical system to be used for representing and processing information behaves quantum mechanically and the type of information is quantum. The question of how classical information can be protected in quantum systems is also interesting but will not be discussed here. We illustrate the principles of quantum error correction by considering quantum versions of the three examples given in "Concepts and Examples" and then add a uniquely quantum example with potentially practical applications in, for example, quantum dot technologies. For an explanation of the basic quantum-information concepts and conventions, see the article "Quantum Information Processing" on page 2.

Trivial Two-Qubit Example. A quantum version of the two-bit example from the previous section consists of two physical qubits, where the errors randomly apply the identity or one of the Pauli operators to the first qubit. The Pauli operators are defined by

$$\mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \text{ and } \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad (7)$$

Explicitly, the errors have the effect

$$|\psi\rangle_{12} \begin{cases} \mathbb{1}|\psi\rangle_{12} & \text{Probability } .25 \\ \sigma_x^{(1)}|\psi\rangle_{12} & \text{Probability } .25 \\ \sigma_y^{(1)}|\psi\rangle_{12} & \text{Probability } .25 \\ \sigma_z^{(1)}|\psi\rangle_{12} & \text{Probability } .25 \end{cases} , \tag{8}$$

where the superscripts in parentheses specify the qubit that an operator acts on. This error model is called completely depolarizing on qubit 1. Obviously, a one-qubit state can be stored in the second physical qubit without being affected by the errors. An encoding operation that implements this observation is

$$|\psi\rangle \rightarrow |0\rangle_1 |\psi\rangle_2 , \tag{9}$$

which realizes an ideal qubit as a two-dimensional subspace of the physical qubits. This subspace is the quantum code for this encoding. To decode, one can discard physical qubit 1 and return qubit 2, which is considered a natural subsystem of the physical system. In this case, the identification of syndrome and information-carrying subsystems is the obvious one associated with the two physical qubits.

Quantum Repetition Code. The repetition code can be used to protect quantum information in the presence of a restricted error model. Let the physical system consist of three qubits. Errors act by independently applying, to each qubit, the flip operator σ_x with probability .25. The classical code can be made into a quantum code by the superposition principle. Encoding one qubit is accomplished by

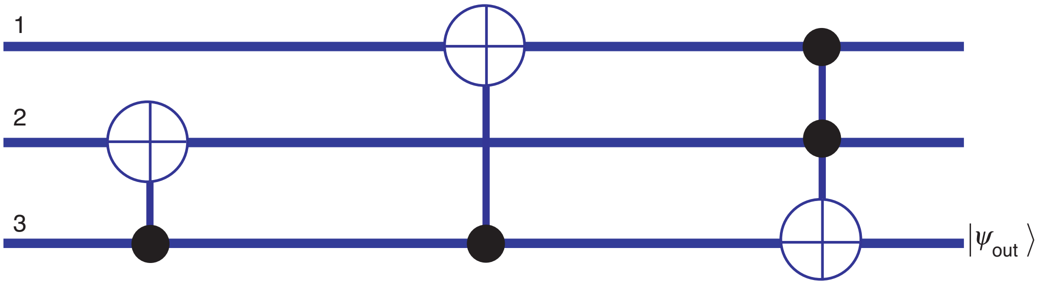
$$\alpha|0\rangle + \beta|1\rangle \rightarrow \alpha|000\rangle + \beta|111\rangle \quad . \tag{10}$$

The associated quantum code is the range of the encoding, that is, the two-dimensional subspace spanned by the encoded states $|000\rangle$ and $|111\rangle$.

As in the classical case, decoding is accomplished by majority logic. However, it must be implemented carefully to avoid destroying quantum coherence in the stored information. One way to do that is to use only unitary operations to transfer the stored information to the output qubit. Figure 5 shows a quantum network that accomplishes this task.

As shown, the decoding network establishes an identification between the three physical qubits and a pair of subsystems consisting of two qubits representing the syndrome subsystem and one qubit for the information-carrying subsystem. On the left side of the correspondence, the information-carrying subsystem is not identifiable with any one (or two) of the physical qubits. Nevertheless, it exists there through the identification.

To obtain a network for encoding, we reverse the decoding network and initialize qubits 2 and 3 in the state $|00\rangle$. The initialization renders the Toffoli gate unnecessary. The complete system with a typical error is shown in Figure 6.



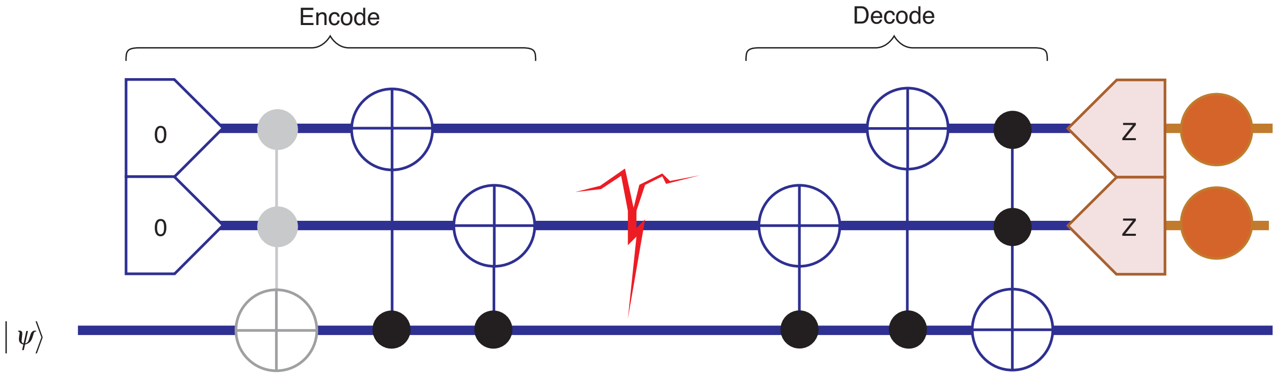
$\lvert 000\rangle$ $\lvert 001\rangle$ $\lvert 010\rangle$ $\lvert 100\rangle$	$\lvert 000\rangle$ $\lvert 011\rangle$ $\lvert 010\rangle$ $\lvert 100\rangle$	$\lvert 000\rangle$ $\lvert 111\rangle$ $\lvert 010\rangle$ $\lvert 100\rangle$	$\lvert 00\rangle\lvert 0\rangle$ $\lvert 11\rangle\lvert 0\rangle$ $\lvert 01\rangle\lvert 0\rangle$ $\lvert 10\rangle\lvert 0\rangle$
$\lvert 111\rangle$ $\lvert 110\rangle$ $\lvert 101\rangle$ $\lvert 011\rangle$	$\lvert 101\rangle$ $\lvert 110\rangle$ $\lvert 111\rangle$ $\lvert 001\rangle$	$\lvert 001\rangle$ $\lvert 110\rangle$ $\lvert 011\rangle$ $\lvert 101\rangle$	$\lvert 00\rangle\lvert 1\rangle$ $\lvert 11\rangle\lvert 1\rangle$ $\lvert 01\rangle\lvert 1\rangle$ $\lvert 10\rangle\lvert 1\rangle$

Figure 5. Majority Logic Decoding into the Output Qubit 3
The effect of the quantum network on the basis states is shown. The top half shows the states with majority 0. The decoded qubit is separated in the last step. The conventions for illustrating quantum networks are explained in the article "Quantum Information Processing" on page 2.

As in the case of the classical repetition code, we can protect against cumulative errors without explicitly decoding and then reencoding, which would cause a temporary loss of protection. Instead, one can find a means for directly resetting the syndrome subsystem to $|00\rangle$ (thus returning the information to the code) before the errors happen again. After resetting in this way, the errors in the correctable set have no effect on the encoded information because they act only on the syndrome subsystem.

Part of the task of designing error-correcting systems is to determine how well the system performs. An important performance measure is the probability of error. In quantum systems, the probability of error is intuitively interpreted as the maximum probability with which we can see a result different from the expected one in any measurement. Specifically, to determine the error, one compares the output $|\psi_o\rangle$ of the system with the input $|\psi\rangle$. An upper bound is obtained if the output is written as a combination of the input state and an error state. For quantum information, combinations are linear combinations (that is, superpositions). Thus $|\psi_o\rangle = \gamma\,|\psi\rangle + |e\rangle$ (see Figure 7). The probability of error is bounded by $\varepsilon = \||e\rangle|^2$ (which we call an error estimate). In general, there are many different ways of writing the output as a combination of an acceptable state and an error term. One attempts to choose the combination that minimizes the error estimate. This choice yields the number ε for which $1 - \varepsilon$ is called fidelity. A fidelity of 1 means that the output is the same (up to a phase factor) as the input.

To illustrate error analysis, we calculate the error for the repetition code example for the two initial states $|0\rangle$ and $(1/\sqrt{2})(|0\rangle + |1\rangle)$.



$$\alpha|0\rangle + \beta|1\rangle \rightarrow \left\{ \begin{matrix} \alpha|00\rangle|0\rangle & \alpha|000\rangle & \alpha|010\rangle & \alpha|01\rangle|0\rangle \\ + & + & + & + \\ \beta|00\rangle|1\rangle & \beta|111\rangle & \beta|101\rangle & \beta|01\rangle|1\rangle \end{matrix} \right\} = |01\rangle\big(\alpha|0\rangle + \beta|1\rangle\big)$$

Figure 6. Networks for the Quantum Repetition Code with a Typical Error
The error that occurred can be determined from the state of the syndrome subsystem, which consists of the top two qubits. The encoding is shown as the reverse of the decoding, starting with an initialized syndrome subsystem. When the decoding is reversed to yield the encoding, there is an initial Toffoli gate (shown in gray). Because of the initialization, this gate has no effect and is therefore omitted in an implementation.

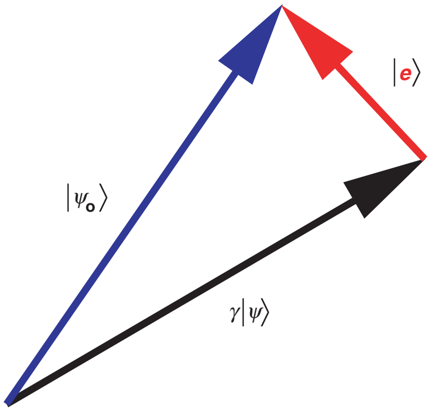


Figure 7. Error Estimate
Any decomposition of the output state $|\psi_o\rangle$ into a "good" state $\gamma|\psi\rangle$ and an (unnormalized) error term $|e\rangle$ gives an estimate $\varepsilon = \||e\rangle|^2$. For pure states, the optimum estimate is obtained when the error term is orthogonal to the input state. To obtain an error estimate for mixtures, one can use any representation of the state as a probabilistic combination of pure states and calculate the probabilistic sum of the pure-state errors.

$$|0\rangle \xrightarrow{\text{encode}} |000\rangle \tag{11}$$

$$\longrightarrow \begin{cases} .75^3 : |000\rangle\ , \\ .25 * .75^2 : |100\rangle\ , \\ .25 * .75^2 : |010\rangle\ , \\ .25 * .75^2 : |001\rangle\ , \\ .25^2 * .75 : |110\rangle\ , \\ .25^2 * .75 : |101\rangle\ , \\ .25^2 * .75 : |011\rangle\ , \\ .25^3 : |111\rangle\ . \end{cases} \tag{12}$$

$$\xrightarrow{\text{decode}} \begin{cases} .4219 : |00\rangle \cdot |0\rangle\ , \\ .1406 : |10\rangle \cdot |0\rangle\ , \\ .1406 : |01\rangle \cdot |0\rangle\ , \\ .1406 : |11\rangle \cdot |0\rangle\ , \\ .0469 : |11\rangle \cdot |1\rangle\ , \\ .0469 : |01\rangle \cdot |1\rangle\ , \\ .0469 : |10\rangle \cdot |1\rangle\ , \\ .0156 : |00\rangle \cdot |1\rangle\ . \end{cases} \tag{13}$$

The final state is a mixture consisting of four correctly decoded components and four incorrectly decoded ones. The probability of each state in the mixture is shown before the colon. The incorrectly decoded information is orthogonal to the encoded information, and its probability is 0.1563, an improvement over the one-qubit error probability of 0.25. The second state behaves quite differently:

$$\frac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big) \xrightarrow{\text{encode}} \frac{1}{\sqrt{2}}\big(|000\rangle + |111\rangle\big) \tag{14}$$

$$\longrightarrow \begin{cases} \vdots \\ .25^2 * .75 : \dfrac{1}{\sqrt{2}}\big(|110\rangle + |001\rangle\big) \\ \vdots \end{cases} \tag{15}$$

$$\xrightarrow{\text{decode}} \begin{cases} \vdots \\ .0469 : \dfrac{1}{\sqrt{2}} |11\rangle \cdot \big(|1\rangle + |0\rangle\big) \\ \vdots \end{cases} \ . \tag{16}$$

Not all error events have been shown, but in each case it can be seen that the state is decoded correctly, so the error is 0. This shows that the error probability can depend significantly on the initial state. To remove this dependence and give a state independent error quantity, one can use the worst-case, the average, or the entanglement error. See the section “Quantum Error Analysis” on page 209.

Quantum Code for a Cyclic System. The shift operators introduced earlier act as permutations of the seven states of the cyclic system. They can therefore be extended to unitary operators on a seven-state cyclic quantum system with logical basis $|0\rangle$, $|1\rangle$, $|2\rangle$, $|3\rangle$, $|4\rangle$, $|5\rangle$, and $|6\rangle$. The error model introduced earlier makes sense here without modification, as does the encoding. The subsystem identification now takes the six-dimensional subspace spanned by $|0\rangle$,.... $|5\rangle$ to a pair consisting of a three-state system with basis $|-1\rangle$, $|0\rangle$, $|1\rangle$ and a qubit. The identification of Equation (6) extends linearly to a unitary subsystem identification. The procedure for decoding is modified as follows: First, a measurement determines whether the state is in the six-dimensional subspace or not. If it is, the identification is used to extract the qubit. Here is an outline of what happens when the state $(1/\sqrt{2})(|0\rangle + |1\rangle)$ is encoded:

$$\frac{1}{\sqrt{2}}\big(|0\rangle+|1\rangle\big) \xrightarrow{\text{encode}} \frac{1}{\sqrt{2}}\big(|1\rangle+|4\rangle\big) \tag{17}$$

$$\longrightarrow \begin{cases} & \vdots \\ .05641e^{-4} & : \frac{1}{\sqrt{2}}\big(|3\rangle+|6\rangle\big) \\ & \vdots \end{cases} \tag{18}$$

$$\xrightarrow{\text{detect}} \begin{cases} .001 : \begin{cases} .5 : & \begin{matrix}\vdots \\ \text{fail} \\ \vdots\end{matrix} \\ .5 : & |3\rangle \end{cases} \end{cases} \tag{19}$$

$$\xrightarrow{\text{decode}} \begin{cases} \quad\vdots \\ .0005 : \ \text{fail} \\ .0005 : \ |-1\rangle \cdot |1\rangle \\ \quad\vdots \end{cases} \tag{20}$$

$$= \begin{cases} \quad\vdots \\ .0005 : \ \text{fail} \\ .0005 : \ |-1\rangle \cdot \left(\frac{1}{2}\big(|0\rangle+|1\rangle\big)+\frac{1}{2}\big(-|0\rangle+|1\rangle\big)\right) \\ \quad\vdots \end{cases} . \tag{21}$$

A “good” state was separated from the output in the case that is shown. The leftover error term has probability amplitude $.0005 * ((1/2)^2 + (1/2)^2) = .00025$, which contributes to the total error (not shown).

Three Quantum Spin-1/2 Particles. Quantum physics provides a rich source of systems with many opportunities for representing and protecting quantum information. Sometimes, it is possible to encode information in such a way that it is protected from the errors indefinitely, without intervention. An example is the trivial two-qubit system discussed before. Whenever error protection without intervention is possible, there is an information-carrying subsystem such that errors act only on the associated syndrome subsystem regardless of the current state. An information-carrying subsystem with this property is called “noiseless.” A physically motivated example of a one-qubit noiseless subsystem can be found in three spin-1/2 particles with errors due to random fluctuations in an external field.

A spin-1/2 particle's state space is spanned by two states, $|\uparrow\rangle$ and $|\downarrow\rangle$. Intuitively, these states correspond to the spin pointing "up" ($|\uparrow\rangle$) or "down" ($|\downarrow\rangle$) in some chosen reference frame. The state space is therefore the same as that of a qubit, and we can make the identifications $|\uparrow\rangle \leftrightarrow |0\rangle$ and $|\downarrow\rangle \leftrightarrow |1\rangle$. An external field causes the spin to rotate according to an evolution of the form

$$|\psi_t\rangle = e^{-i\left(u_x\sigma_x + u_y\sigma_y + u_z\sigma_z\right)t/2}|\psi\rangle . \tag{22}$$

The vector $\mathbf{u} = (u_x, u_y, u_z)$ characterizes the direction of the field and the strength of the spin's interaction with the field. This situation arises, for example, in nuclear magnetic resonance with spin-1/2 nuclei, where the fields are magnetic fields (see the article "NMR and Quantum Information Processing" on page 226).

Now consider the physical system composed of three spin-1/2 particles with errors acting as identical rotations of the three particles. Such errors occur if they are due to a uniform external field that fluctuates randomly in direction and strength. The evolution caused by a uniform field is given by

$$\begin{aligned} |\psi_t\rangle_{123} &= e^{-i\left(u_x\sigma_x^{(1)} + u_y\sigma_y^{(1)} + u_z\sigma_z^{(1)}\right)t/2} e^{-i\left(u_x\sigma_x^{(2)} + u_y\sigma_y^{(2)} + u_z\sigma_z^{(2)}\right)t/2} e^{-i\left(u_x\sigma_x^{(3)} + u_y\sigma_y^{(3)} + u_z\sigma_z^{(3)}\right)t/2}|\psi\rangle_{123} \\ &= e^{-i\left(u_x\left(\sigma_x^{(1)}+\sigma_x^{(2)}+\sigma_x^{(3)}\right) + u_y\left(\sigma_y^{(1)}+\sigma_y^{(2)}+\sigma_y^{(3)}\right) + u_z\left(\sigma_z^{(1)}+\sigma_z^{(2)}+\sigma_z^{(3)}\right)\right)t/2}|\psi\rangle_{123} \\ &= e^{-i\left(u_xJ_x + u_yJ_y + u_zJ_z\right)t}|\psi\rangle_{123} , \end{aligned} \tag{23}$$

with $J_u = (\sigma_u^{(1)} + \sigma_u^{(2)} + \sigma_u^{(3)})/2$ for $u = x$, y, and z. We can exhibit the error operators arising from a uniform field in a compact form by defining $\boldsymbol{J} = (J_x, J_y, J_z)$ and $\boldsymbol{v} = (u_x, u_y, u_z)t$. Then the error operators are given by $E(\boldsymbol{v}) = e^{-i\boldsymbol{v}\cdot\boldsymbol{J}}$, where the dot product in the exponent is calculated like the standard vector dot product.

For a one-qubit noiseless subsystem, the key property of the error model is that the errors are symmetric under any permutation of the three particles. A permutation of the particles acts on the particles' state space by permuting the labels in the logical states. For example, the permutation π that swaps the first two particles acts on logical states as

$$\pi|a\rangle_1|b\rangle_2|c\rangle_3 = |a\rangle_2|b\rangle_1|c\rangle_3 = |b\rangle_1|a\rangle_2|c\rangle_3 . \tag{24}$$

To say that the errors are symmetric under particle permutations means that each error E satisfies $\pi^{-1}E\pi = E$, or equivalently, $E\pi = \pi E$ (E commutes with π). To see that this condition is satisfied, write

$$\begin{aligned}\pi^{-1}E(\boldsymbol{\nu})\pi &= \pi^{-1}e^{-i\boldsymbol{\nu}\cdot\boldsymbol{J}}\pi \\ &= e^{-i\pi^{-1}(\boldsymbol{\nu}\cdot\boldsymbol{J})\pi} \\ &= e^{-i\boldsymbol{\nu}\left(\pi^{-1}\boldsymbol{J}\pi\right)} \quad . \end{aligned} \tag{25}$$

If π permutes particle a to particle b, then $\pi^{-1}\sigma_u^{(a)}\pi = \sigma_u^{(b)}$. It follows that $\pi^{-1}\boldsymbol{J}\pi = \boldsymbol{J}$. This expression shows that the errors commute with the particle permutations and therefore cannot distinguish between the particles. An error model satisfying this property is called a collective error model.

If a noiseless subsystem exists, then learning the symmetries of the error model suffices for constructing the subsystem. This procedure is explained later, in "Conserved Quantities, Symmetries, and Noiseless Subsystems." For the three spin-1/2 system, the procedure results in a one-qubit noiseless subsystem protected from all collective errors. We first exhibit the subsystem identification and then discuss its properties to explain why it is noiseless. As in the case of the seven-state cyclic system, the identification involves a proper subspace of the physical system's state space. The subsystem identification involves a four-dimensional subspace and is defined by the following correspondence:

$$\begin{aligned}
&\frac{1}{\sqrt{3}}\left(|\downarrow\rangle_1|\uparrow\rangle_2|\uparrow\rangle_3 + e^{-i2\pi/3}|\uparrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3 + e^{i2\pi/3}|\uparrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3\right) \leftrightarrow |\uparrow\rangle\cdot|\mathsf{0}\rangle \\
&\frac{1}{\sqrt{3}}\left(|\downarrow\rangle_1|\uparrow\rangle_2|\uparrow\rangle_3 + e^{i2\pi/3}|\uparrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3 + e^{-i2\pi/3}|\uparrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3\right) \leftrightarrow |\uparrow\rangle\cdot|\mathsf{1}\rangle \\
-&\frac{1}{\sqrt{3}}\left(|\uparrow\rangle_1|\downarrow\rangle_2|\downarrow\rangle_3 + e^{-i2\pi/3}|\downarrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3 + e^{i2\pi/3}|\downarrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3\right) \leftrightarrow |\downarrow\rangle\cdot|\mathsf{0}\rangle \\
-&\frac{1}{\sqrt{3}}\left(|\uparrow\rangle_1|\downarrow\rangle_2|\downarrow\rangle_3 + e^{i2\pi/3}|\downarrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3 + e^{-i2\pi/3}|\downarrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3\right) \leftrightarrow |\downarrow\rangle\cdot|\mathsf{1}\rangle \quad .
\end{aligned} \tag{26}$$

The state labels for the syndrome subsystem (before the dot in the expressions on the right side) identify it as a spin-1/2 subsystem. In particular, it responds to the errors caused by uniform fields in the same way as the physical spin-1/2 particles. This behavior is caused by $2J_u$ acting as the u-Pauli operator on the syndrome subsystem. To confirm this property, we apply $2J_u$ to the logical states of Equation (26) for $u = z, x$. The property for $u = y$ then follows because $i\sigma_y = \sigma_z\sigma_x$. Consider $2J_z$. Each of the four states shown in Equation (26) is an eigenstate of $2J_z$. For example, the physical state for $|\uparrow\rangle \cdot |\mathsf{0}\rangle$ is a superposition of states with two spins up ($\uparrow$) and one spin down ($\downarrow$). The eigenvalue of such a state with respect to $2J_z$ is the difference Δ between the number of spins that are up and down. Thus, $2J_z|\uparrow\rangle \cdot |\mathsf{0}\rangle = |\uparrow\rangle \cdot |\mathsf{0}\rangle$. The difference is also $\Delta = 1$ for $|\uparrow\rangle \cdot |\mathsf{1}\rangle$ and $\Delta = -1$ for $|\downarrow\rangle \cdot |\mathsf{0}\rangle$ and $|\downarrow\rangle \cdot |\mathsf{1}\rangle$. Therefore, $2J_z$ acts as the z-Pauli operator on the syndrome subsystem. To confirm this behavior for $2J_x$, we compute $2J_x|\uparrow\rangle \cdot |\mathsf{0}\rangle$.

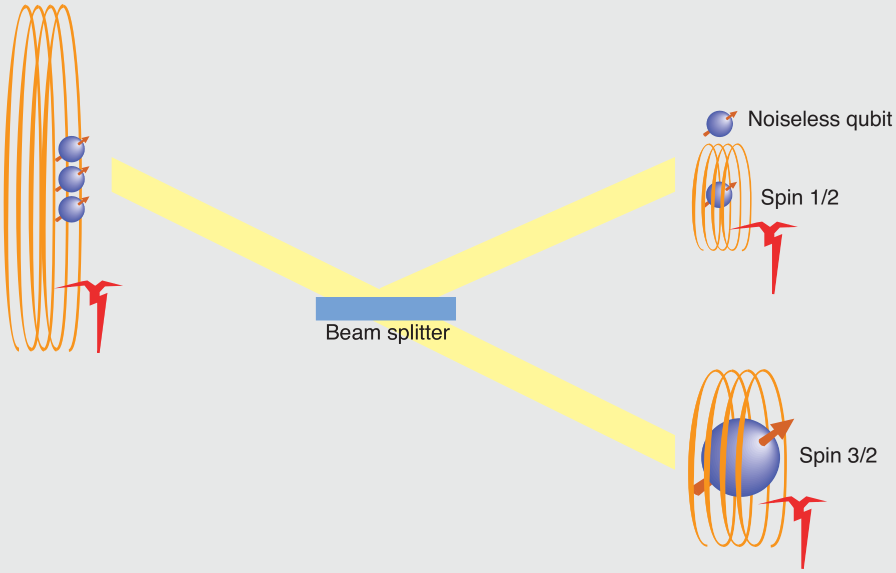
$$
\begin{aligned}
2J_x\left|\uparrow\right\rangle\cdot\left|\mathsf{o}\right\rangle = 2J_x\frac{1}{\sqrt{3}}\Big(&\left|\downarrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\uparrow\right\rangle_3+e^{-i2\pi/3}\left|\uparrow\right\rangle_1\left|\downarrow\right\rangle_2\left|\uparrow\right\rangle_3+e^{i2\pi/3}\left|\uparrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\downarrow\right\rangle_3\Big)\\
= \quad & \frac{1}{\sqrt{3}}\left(\sigma_x^{(1)}+\sigma_x^{(2)}+\sigma_x^{(3)}\right)\left|\downarrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\uparrow\right\rangle_3\\
& +e^{-i2\pi/3}\frac{1}{\sqrt{3}}\left(\sigma_x^{(1)}+\sigma_x^{(2)}+\sigma_x^{(3)}\right)\left|\uparrow\right\rangle_1\left|\downarrow\right\rangle_2\left|\uparrow\right\rangle_3\\
& +e^{i2\pi/3}\frac{1}{\sqrt{3}}\left(\sigma_x^{(1)}+\sigma_x^{(2)}+\sigma_x^{(3)}\right)\left|\uparrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\downarrow\right\rangle_3\\
= \quad & \frac{1}{\sqrt{3}}\Big(\left|\uparrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\uparrow\right\rangle_3+\left|\downarrow\right\rangle_1\left|\downarrow\right\rangle_2\left|\uparrow\right\rangle_3+\left|\downarrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\downarrow\right\rangle_3\Big)\\
& +e^{-i2\pi/3}\frac{1}{\sqrt{3}}\Big(\left|\downarrow\right\rangle_1\left|\downarrow\right\rangle_2\left|\uparrow\right\rangle_3+\left|\uparrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\uparrow\right\rangle_3+\left|\uparrow\right\rangle_1\left|\downarrow\right\rangle_2\left|\downarrow\right\rangle_3\Big)\\
& +\ e^{i2\pi/3}\frac{1}{\sqrt{3}}\Big(\left|\downarrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\downarrow\right\rangle_3+\left|\uparrow\right\rangle_1\left|\downarrow\right\rangle_2\left|\downarrow\right\rangle_3+\left|\uparrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\uparrow\right\rangle_3\Big)\\
= \quad & \frac{1}{\sqrt{3}}\left(1+e^{-i2\pi/3}+e^{i2\pi/3}\right)\left|\uparrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\uparrow\right\rangle_3\\
& +\frac{1}{\sqrt{3}}\left(e^{-i2\pi/3}+e^{i2\pi/3}\right)\left|\uparrow\right\rangle_1\left|\downarrow\right\rangle_2\left|\downarrow\right\rangle_3\\
& +\frac{1}{\sqrt{3}}\left(1+\phantom{e^{-i2\pi/3}+}e^{i2\pi/3}\right)\left|\downarrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\downarrow\right\rangle_3\\
& +\frac{1}{\sqrt{3}}\left(1+e^{-i2\pi/3}\phantom{+e^{i2\pi/3}}\right)\left|\downarrow\right\rangle_1\left|\downarrow\right\rangle_2\left|\uparrow\right\rangle_3\\
= \quad & -\frac{1}{\sqrt{3}}\Big(\left|\uparrow\right\rangle_1\left|\downarrow\right\rangle_2\left|\downarrow\right\rangle_3+e^{-i2\pi/3}\left|\downarrow\right\rangle_1\left|\uparrow\right\rangle_2\left|\downarrow\right\rangle_3+e^{i2\pi/3}\left|\downarrow\right\rangle_1\left|\downarrow\right\rangle_2\left|\uparrow\right\rangle_3\Big)\\
= \quad & \left|\downarrow\right\rangle\cdot\left|\mathsf{o}\right\rangle \quad . \qquad (27)
\end{aligned}
$$

Similarly, one can check that, for the other logical states, the effect of $2J_x$ is to flip the orientation of the syndrome spin. That the subsystem identified in Equation (26) is noiseless now follows from the fact that the errors $E(\boldsymbol{v})$ are exponentials of sums of the syndrome spin operators J_u. The errors therefore act as the identity on the information-carrying subsystem.

The noiseless qubit supported by three spin-1/2 particles with collective errors is another example in which the subsystem identification does not involve the whole state space of the system. In this case, the errors of the error model cannot remove amplitude from the subspace. As a result, if we detect an error, that is, if we find that the system's state is in the orthogonal complement of the subspace of the subsystem identification, we can deduce that either the error model is inadequate or we introduced errors in the manipulations required for transferring information to the noiseless qubit.

The noiseless subsystem of three spin-1/2 particles can be physically motivated by an analysis of quantum spin numbers. This analysis is outlined in the box on the opposite page.

Creating a Noiseless Subsystem from Three Spin-1/2 Particles



The left side shows the three particles with errors caused by fluctuations in a uniform magnetic field depicted by a noisy coil. The spin along direction u ($u = x, y, z$) can be measured, and its expectation is given by $\langle\psi|J_u|\psi\rangle$, where $|\psi\rangle$ is the quantum state of the particles and J_u is the total spin observable along the u-axis given by the half sum of the u-Pauli matrices of the particles as defined in the text. The squared magnitude of the total spin is given by the expectation of the observable $J^2 = \boldsymbol{J} \cdot \boldsymbol{J} = J_x^2 + J_y^2 + J_z^2$. The observable J^2 commutes with the J_u and therefore also with the errors $E(\boldsymbol{v}) = e^{-i\boldsymbol{v}\cdot\boldsymbol{J}}$ caused by uniform field fluctuations. This statement can be verified directly, or one can note that $E(\boldsymbol{v})$ acts on $\boldsymbol{J}$ as a rotation in three dimensions, and as one would expect, such rotations preserve the squared length J^2 of $\boldsymbol{J}$. It now follows that the eigenspaces of J^2 are invariant under the errors and, therefore, that the eigenspaces are good places to look for noiseless subsystems. The eigenvalues of J^2 are of the form $j\,(j + 1)$, where j is the spin quantum number of the corresponding eigenspace. There are two eigenspaces, one with spin $j = 1/2$ and the other with spin $j = 3/2$.

The figure shows a thought experiment that involves passing the three-particle system through a type of beam splitter or Stern-Gerlach apparatus sensitive to J^2. Using such a beam splitter, the system of particles can be made to go in one of two directions, depending on j. In the figure, if the system's state is in the spin-3/2 subspace, it passes through the beam splitter; if it is in the spin-1/2 subspace, the system is reflected up. It can be shown that the subspace with $j = 3/2$ is four dimensional and spanned by the states that are symmetric under particle permutations. Unfortunately, there is no noiseless subsystem in this subspace (refer to the section "Conserved Quantities, Symmetries, and Noiseless Subsystems"). The spin-1/2 subspace is also four dimensional and spanned by the states in Equation (26). The spin-1/2 property of the subspace implies that the spin operators J_u act in a way that is algebraically identical to the way $\sigma_u/2$ acts on a single spin-1/2 particle. This property implies the existence of the syndrome subsystem introduced in the text. Conventionally, the spin-1/2 subspace is thought of as consisting of two orthogonal two-dimensional subspaces, each behaving like a spin-1/2 with respect to the J_u. This choice of subspaces is not unique, but by associating them with two logical states of a noiseless qubit, one can obtain the subsystem identification of Equation (26). Some care needs to be taken to ensure that the noiseless qubit operators commute with the J_u, as they should. In the thought experiment shown in the figure, one can imagine unitarily rotating the system emerging in the upper path to make explicit the syndrome spin-1/2 subsystem and the noiseless qubit with which it must be paired. The result of this rotation is shown.

Error Models

We have seen several models of physical systems and errors in the examples of the previous sections. Most physical systems under consideration for QIP consist of particles or degrees of freedom that are spatially localized, a feature reflected in the error models that are usually investigated. Because we also expect the physically realized qubits to be localized, the standard error models deal with quantum errors that act independently on different qubits. Logically realized qubits, such as those implemented by subsystems different from the physically obvious ones, may have more complicated residual-error behaviors.

The Standard Error Models for Qubits. The most investigated error model for qubits consists of independent, depolarizing errors. This model has the effect of completely depolarizing each qubit independently with probability p—see Equation (8). For one qubit, the model is the least biased in the sense that it is symmetric under rotations. As a result, every state of the qubit is equally affected. Independent depolarizing errors are considered to be the quantum analogue of the classical independent bit-flip error model.

Depolarizing errors are not typical for physically realized qubits. However, given the ability to control individual qubits, it is possible to enforce the depolarizing model (see below). Consequently, error correction methods designed to control depolarizing errors apply to all independent error models. Nevertheless, it is worth keeping in mind that given detailed knowledge of the physical errors, a special purpose method is usually better than one designed for depolarizing errors. We therefore begin by showing how one can think about arbitrary error models.

There are several different ways of describing errors affecting a physical system (or “sys” for short) of interest. For most situations, in particular if the initial state of the system is pure, errors can be thought of as being the result of coupling to an initially independent environment for some time. Because of this coupling, the effect of error can always be represented by the process of adjoining an environment (or “env” for short) in some initial state $|0\rangle_{\text{env}}$ to the arbitrary state $|\psi\rangle_{\text{sys}}$, followed by a unitary coupling evolution $U^{(\text{env, sys})}$ acting jointly on the environment and the system. Symbolically, the process can be written as the map

$$|\psi\rangle_{\text{sys}} \rightarrow U^{(\text{env, sys})}|0\rangle_{\text{env}}|\psi\rangle_{\text{sys}} \quad . \tag{28}$$

Choosing an arbitrary orthonormal basis consisting of the states $|e\rangle_{\text{env}}$ for the state space of the environment, the process can be rewritten in the form

$$\begin{aligned} |\psi\rangle_{\text{sys}} &\rightarrow \mathbb{1}^{(\text{env})}\, U^{(\text{env,sys})}|0\rangle_{\text{env}}|\psi\rangle_{\text{sys}} \\ &= \left(\sum_e |e\rangle_{\text{env}}\, {}^{\text{env}}\langle e|\right) U^{(\text{env,sys})}|0\rangle_{\text{env}}|\psi\rangle_{\text{sys}} \\ &= \sum_e |e\rangle_{\text{env}} \left({}^{\text{env}}\langle e|U^{(\text{env,sys})}|0\rangle_{\text{env}}\right)|\psi\rangle_{\text{sys}} \\ &= \sum_e |e\rangle_{\text{env}} A_e^{(\text{sys})}|\psi\rangle_{\text{sys}} \ , \end{aligned} \tag{29}$$

where the last step defines operators $A_e^{(\mathrm{sys})}$ acting on the physical system by $A_e^{(\mathrm{sys})} = {}^{\mathrm{env}}\langle e|U^{(\mathrm{env,\,sys})}|0\rangle_{\mathrm{env}}$. The expression $\sum_e |e\rangle_{\mathrm{env}} A_e^{(\mathrm{sys})}$ is called an environment-labeled operator. The unitarity condition implies that $\sum_e A_e^\dagger A_e = \mathbb{1}$ (with system labels omitted). The environment basis $|e\rangle_{\mathrm{env}}$ need not represent any physically meaningful choice of basis of a real environment. For error analysis, the states $|e\rangle_{\mathrm{env}}$ are formal states that label the error operators A_e. One can use an expression of the form shown in Equation (29) even when the $|e\rangle$ are not normalized or orthogonal, keeping in mind that, as a result, the identity implied by the unitarity condition changes.

Note that the state on the right side of Equation (29), representing the effect of the errors, is correlated with the environment. This means that after removing (or "tracing over") the environment, the state of the physical system is usually mixed. Instead of introducing an artificial environment, we can also describe the errors by using the density operator formalism for mixed states. Define $\rho = |\psi\rangle_{\mathrm{sys}}{}^{\mathrm{sys}}\langle\psi|$. The effect of the errors on the density matrix ρ is given by the transformation

$$\rho \rightarrow \sum_e A_e \rho A_e^\dagger \ . \tag{30}$$

This is the "operator sum" formalism (Kraus 1983).

The two ways of writing the effects of errors can be applied to the depolarizing-error model for one qubit. As an environment-labeled operator, depolarization with probability p can be written as

$$\sqrt{1-p}\,|0\rangle_{\mathrm{env}}\mathbb{1} + \frac{\sqrt{p}}{2}\left(|1\rangle_{\mathrm{env}}\mathbb{1} + |x\rangle_{\mathrm{env}}\sigma_x + |y\rangle_{\mathrm{env}}\sigma_y + |z\rangle_{\mathrm{env}}\sigma_z\right) \ , \tag{31}$$

where we introduced five abstract, orthonormal environment states to label the different events. In this case, one can think of the model as applying no error with probability $1 - p$ or completely depolarizing the qubit with probability p. The latter event is represented by applying one of $\mathbb{1}$, σ_x, σ_y, or σ_z with equal probability $p/4$. To be able to think of the model as randomly applied Pauli matrices, it is crucial that the environment states labeling the different Pauli matrices be orthogonal. The square roots of the probabilities appear in the operator because, in an environment-labeled operator, it is necessary to give quantum amplitudes. Environment-labeled operators are useful primarily because of their great flexibility and redundancy.

In the operator sum formalism, depolarization with probability p transforms the input density matrix ρ as

$$\begin{aligned} \rho &\rightarrow (1-p)\rho + \frac{p}{4}\left(\mathbb{1}\,\rho\mathbb{1} + \sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z\right) \\ &= (1-3p/4)\rho + \frac{p}{4}\left(\sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z\right) \ . \end{aligned} \tag{32}$$

Because the operator sum formalism has less redundancy, it is easier to tell when two error effects are equivalent.

In the remainder of this section, we discuss how one can use active intervention to simplify the error model. To realize this simplification, we intentionally randomize the qubit so that the environment cannot distinguish between the different axes defined by the Pauli spin matrices. Here is a simple randomization that actively converts an arbitrary error model for a qubit into one that consists of randomly applying Pauli operators according to some distribution. The distribution is not necessarily uniform, so the new error model is not yet depolarizing. Before the errors act, apply a random Pauli operator σ_u ($u = 0, x, y, z$, $\sigma_0 = \mathbb{1}$). After the errors act, apply the inverse of that operator, $\sigma_u^{-1} = \sigma_u$; then "forget" which operator was applied. This randomization method is called twirling (Bennett et al. 1996). To understand twirling, we use environment-labeled operators to demonstrate some of the techniques useful in this context. The sequence of actions implementing twirling can be written as follows (omitting labels for the physical system):

$$\begin{aligned}
|\psi\rangle &\to \frac{1}{2}\sum\nolimits_u |u\rangle_{\mathrm{C}}\sigma_u|\psi\rangle && \text{Apply a random } \sigma_u \text{ remembering } u \text{ with the help of the system } C. \\
&\to \sum\nolimits_e |e\rangle_{\mathrm{env}} \frac{1}{2}\sum\nolimits_u |u\rangle_{\mathrm{C}} A_e\sigma_u|\psi\rangle && \text{Errors act.} \\
&\to \sum\nolimits_e |e\rangle_{\mathrm{env}} \frac{1}{2}\sum\nolimits_u |u\rangle_{\mathrm{C}} \sigma_u A_e\sigma_u|\psi\rangle && \text{Apply } \sigma_u = \sigma_u^{-1}. \\
&\to \sum\nolimits_{eu} |eu\rangle_{\mathrm{env,C}} \frac{1}{2}\sigma_u A_e\sigma_u|\psi\rangle \quad . && \text{Forget which } u \text{ was used by absorbing its memory in the environment.}
\end{aligned} \tag{33}$$

The system C that was artificially introduced to carry the memory of u may be a classical memory because there is no need for coherence between different $|u\rangle_{\mathrm{C}}$.

To determine the equivalent random Pauli operator error model, it is necessary to rewrite the total effect of the procedure using an environment-labeled sum involving orthogonal environment states and Pauli operators. To do so, express A_e as a sum of the Pauli operators, $A_e = \sum_v \alpha_{ev}\sigma_v$, using the fact that the σ_v are a linear basis for the space of one-qubit operators. Recall that σ_u anticommutes with σ_v if $0 \neq u \neq v \neq 0$. Thus, $\sigma_u\,\sigma_v\,\sigma_u = (-1)^{\langle v,u\rangle}\sigma_v$, where $\langle v, u\rangle = 1$ if $0 \neq u \neq v \neq 0$ and $\langle v, u\rangle = 0$ otherwise. We can now rewrite the last expression of Equation (33) as follows:

$$\begin{aligned}
\sum_{eu} |eu\rangle_{\mathrm{env,C}} \frac{1}{2}\sigma_u A_e\sigma_u|\psi\rangle &= \sum_{eu} |eu\rangle_{\mathrm{env,C}} \frac{1}{2}\sigma_u \sum_v \alpha_{ev}\sigma_v\sigma_u|\psi\rangle \\
&= \sum_v \left(\sum_{eu}\frac{1}{2}\alpha_{ev}(-1)^{\langle v,u\rangle}|eu\rangle_{\mathrm{env,C}}\right)\sigma_v|\psi\rangle \quad .
\end{aligned} \tag{34}$$

It can be checked that the states $(1/2)\sum_u(-1)^{\langle v,u\rangle}|eu\rangle_{\mathrm{env,C}}$ are orthonormal for different e and v. As a result, the states $\sum_{eu}(1/2)\alpha_{ev}(-1)^{\langle v,u\rangle}|eu\rangle_{\mathrm{env,C}}$ are orthogonal for different v and have probability (square norm) given by $p_v = \sum_e |\alpha_{ev}|^2$. Introducing $\sqrt{p_v}|\tilde{v}\rangle_{\mathrm{env,C}} = \sum_{eu}(1/2)\alpha_{ev}(-1)^{\langle v,u\rangle}|eu\rangle_{\mathrm{env,C}}$, we can write the sum of Equation (34) as

$$\sum_v \left(\sum_{eu}\frac{1}{2}\alpha_{ev}(-1)^{\langle v,u\rangle}|eu\rangle_{\mathrm{env,C}}\right)\sigma_v|\psi\rangle = \sum_v \sqrt{p_v}\,|\tilde{v}\rangle_{\mathrm{env,C}}\sigma_v|\psi\rangle \quad , \tag{35}$$

showing that the twirled error model behaves like randomly applied Pauli matrices with σ_v applied with probability p_v. It is a recommended exercise to reproduce the above argument using the operator sum formalism.

To obtain the standard depolarizing error model with equal probabilities for the Pauli matrices, it is necessary to strengthen the randomization procedure by applying a random member U of the group generated by the 90° rotations around the x-, y-, and z-axis before the error and then undoing U by applying U^{-1}.

Randomization can be used to transform any one-qubit error model into the depolarizing error model. This explains why the depolarizing model is so useful for analyzing error correction techniques in situations in which errors act independently on different qubits. However, in many physical situations, the independence assumptions are not satisfied. For example, errors from common internal couplings between qubits are generally pairwise correlated to first order. In addition, the operations required to manipulate the qubits and to control the encoded information act on pairs at a time, which tends to spread even single-qubit errors. Still, in all these cases, the primary error processes are local. This means that there usually exists an environment-labeled sum expression for the total error process in which the amplitudes associated with errors acting simultaneously at k locations in time and space decrease exponentially with k. In such cases, error correction methods that handle all or most errors involving sufficiently few qubits are still applicable.

Quantum Error Analysis. One of the most important consequences of the subsystems interpretation of encoding quantum information in a physical system is that the encoded quantum information can be error-free even though errors have severely changed the state of the physical system. Almost trivially, any error operator acting only on the syndrome subsystem has no effect on the quantum information. The goal of error correction is to actively intervene and maintain the syndrome subsystem in states where the dominant error operators continue to have little effect on the information of interest. An important issue in analyzing error correction methods is to estimate the residual error in the encoded information. A simple example of how that can be done was discussed for the quantum repetition code. The same ideas can be applied in general. Let sys be the physical system in which the information is encoded, and $|\psi\rangle_{\text{sys}}$ an initial state containing such information with the syndrome subsystem appropriately prepared. Errors and error-correcting operations modify the state. The new state can be expressed with environment labeling as $\sum_e |e\rangle_{\text{env}} A_e{}^{(\text{sys})} |\psi\rangle_{\text{sys}}$. In view of the partitioning into information-carrying and syndrome subsystems, good states $|e\rangle_{\text{env}}$ are those states for which $A_e{}^{(\text{sys})}$ acts only on the syndrome subsystem, given that the syndrome has been prepared. The remaining states $|e\rangle$ form the set of bad states, B. The error probability p_e can be bounded from above by

$$
\begin{aligned}
p_e &\leq \left| \sum_{e \in B} |e\rangle_{\text{env}} A_e{}^{(\text{sys})} |\psi\rangle_{\text{sys}} \right|^2 \\
&\leq \left(\sum_{e \in B} \left| |e\rangle_{\text{env}} \right| \left| A_e{}^{(\text{sys})} \right|_1 \right)^2 ,
\end{aligned}
\tag{36}
$$

where $|A|_1 = \max_\phi \langle \phi | A | \phi \rangle$, the maximum being taken over normalized states. The second inequality usually leads to a gross overestimate but is independent of the encoded information and often suffices for obtaining good results. Because the environment-labeled sum is not unique, a goal of the representation of the errors acting on the system is to use "good" operators to the largest extent possible. The flexibility of these error expansions makes them very useful for analyzing error models in conjunction with error correction methods.

In principle, we can obtain better expressions for p_e by calculating the density matrix ρ of the state of the subsystem containing the desired quantum information. This calculation involves tracing over the syndrome subsystem. The matrix ρ can then be compared to the intended state. If the intended state is pure, given by $|\phi\rangle$, the probability of error is given by $1 - \langle\phi|\rho|\phi\rangle$, which is the probability that a measurement that distinguishes between $|\phi\rangle$ and its orthogonal complement fails to detect $|\phi\rangle$. The quantity $\langle\phi|\rho|\phi\rangle$ is called the fidelity of the state ρ.

For applications to communication, the goal is to be able to reliably transmit arbitrary states through a communication channel, which may be physical or realized via an encoding/decoding scheme. It is therefore important to characterize the reliability of the channel independent of the information transmitted. Equation (36) can be used to obtain state-independent bounds on the error probability but does not readily provide a single measure of reliability. One way to quantify the reliability is to identify the error of the channel with the average error ε_a over all possible input states. The reliability is then given by the average fidelity $1 - \varepsilon_a$. Another elegant way appropriate for QIP is to use the entanglement fidelity (Schumacher 1996). Entanglement fidelity measures the error when the input is maximally entangled with an identical reference system. In this process, the reference system is imagined to be untouched, so that the state of the reference system, together with the output state, can be compared with the original entangled state. For a one-qubit channel labeled sys, the reference system is a qubit, which we label "ref." An initial, maximally entangled state is

$$|B\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle_{\text{ref}}|0\rangle_{\text{sys}} + |1\rangle_{\text{ref}}|1\rangle_{\text{sys}}\right) \quad . \tag{37}$$

The reference qubit is assumed to be perfectly isolated and not affected by any errors. The final state $\rho^{(\text{ref,sys})}$ is compared with $|B\rangle$, which gives the entanglement fidelity according to the formula $f_e = \langle B|\rho^{(\text{ref,sys})}|B\rangle$. The entanglement error is $\varepsilon_e = 1 - f_e$. It turns out that this definition does not depend on the choice of maximally entangled state. Fortunately, the entanglement error and the average error ε_a are related by a linear expression:

$$\varepsilon_a = \frac{2}{3}\varepsilon_e \quad . \tag{38}$$

For k-qubit channels, the constant 2/3 is replaced by $2^k/(2^k + 1)$. Experimental measurements of these fidelities do not require the reference system. There are simple averaging formulas to express them in terms of the fidelities for transmitting each of a sufficiently large set of pure states. An example of the experimental determination of the entanglement fidelity when the channel is realized by error correction is provided in Knill et al. (2001).

From Quantum Error Detection to Error Correction

In the independent depolarizing error model with small probability p of depolarization, the most likely errors are those that affect a small number of qubits. That is, if we define the weight of a product of Pauli operators to be the number of qubits affected, the dominant errors are those of small weight. Because the probability of a nonidentity Pauli operator is $3p/4$—see Equation (31)—one expects about $(3p/4)n$ of n qubits to be changed. As a result, good error-correcting codes are considered to be those for which all errors of weight $\leq e \cong (3p/4)n$ can be corrected. It is desirable that e have a high rate, which means that it is a large fraction of the total number of qubits n (the length of the code). Combinatorially, good codes are characterized by a high minimum distance, a concept that arises naturally in the context of error detection.

Quantum Error Detection. Let C be a quantum code, that is, a subspace of the state space of a quantum system. Let P be the operator that projects onto C, and $P^{\perp} = \mathbb{1} - P$ the one that projects onto the orthogonal complement. Then the pair P, $P^{\perp}$ is associated with a measurement that can be used to determine whether a state is in the code or not. If the given state is $|\psi\rangle$, the result of the measurement is $P|\psi\rangle$ with probability $|P|\psi\rangle|^2$ and $P^{\perp}|\psi\rangle$ otherwise. As in the classical case, an error-detection scheme consists of preparing the desired state $|\psi_i\rangle \in C$, transmitting it through, say, a quantum channel, then measuring whether the state is still in the code, accepting the state if it is, and rejecting it otherwise. We say that C detects error operator E if states accepted after E had acted are unchanged except for an overall scale. Using the projection operators, this is the statement that for every state $|\psi_i\rangle \in C$, $PE|\psi_i\rangle = \lambda_E |\psi_i\rangle$. Because $P|\psi\rangle$ is in the code for every $|\psi\rangle$, it follows that $PEP|\psi\rangle = \lambda_E P|\psi\rangle$. It follows that a characterization of detectability is given by Theorem 3.

Theorem 3. E is detectable by C if and only if $PEP = \lambda_E P$ for some λ_E.

A second characterization is given by Theorem 4.

Theorem 4. E is detectable by C if and only if for all $|\psi\rangle, |\phi\rangle \in C$, $\langle\psi|E|\phi\rangle = \lambda_E \langle\psi|\phi\rangle$ for some λ_E.

A third characterization is obtained by taking the condition for classical detectability in Theorem 1 and replacing $\neq$ by orthogonal to:

Theorem 5. E is detectable by C if and only if for all $|\phi\rangle$, $|\psi\rangle$ in the code with $|\phi\rangle$ orthogonal to $|\psi\rangle$, $E|\phi\rangle$ is orthogonal to $|\psi\rangle$.

For a given code C, the set of detectable errors is closed under linear combinations. That is, if E_1 and E_2 are both detectable, then so is $\alpha E_1 + \alpha E_2$. This useful property implies that, to check detectability, one has to consider only the elements of a linear basis for the space of errors of interest.

Consider n-qubits with independent depolarizing errors. A robust error-detecting code should detect as many of the small-weight errors as possible. This requirement motivates the definition of minimum distance: The code C has minimum distance d if the smallest-weight product of Pauli operators E for which C does not detect E is d. The notion comes from classical codes for bits, where a set of code words C' has minimum distance d if the smallest number of flips required to change one code word in C' into another one in C' is d. For example, the repetition code for three bits has minimum distance 3. Note that the minimum distance for the quantum repetition code is 1: Applying $\sigma_z^{(1)}$ preserves the code and changes the sign of $|111\rangle$ but not of $|000\rangle$. As a result, $\sigma_z^{(1)}$ is not detectable. The notion of minimum distance can be generalized for error models with specified first-order error operators (Knill et al. 2000). In the case of depolarizing errors, the first-order error operators are single-qubit Pauli matrices, which are the errors of weight 1.

Quantum Error Correction. Let $\mathcal{E} = \{E_0 = \mathbb{1}, E_1, \ldots\}$ be the set of errors that we wish to be able to correct. When a decoding procedure for the code C exists such that all errors in $\mathcal{E}$ are corrected, we say that $\mathcal{E}$ is correctable (by C). A situation in which correctability of $\mathcal{E}$ is apparent occurs when the errors E_i are unitary operators satisfying the condition that E_iC are mutually orthogonal subspaces. The repetition code has this property for the set of errors consisting of the identity and Pauli operators acting on a single qubit. In this situation, the procedure for decoding is to first make a projective measurement and determine which of the subspaces E_iC the state is in and then to apply the inverse of the error operator, that is, $E_i^\dagger$. This situation is not far from the generic one. One characterization of correctability is described in Theorem 6.

Theorem 6. $\mathcal{E}$ is correctable if and only if there is a linear transformation of the set $\mathcal{E}$ such that the operators E_i' in the new set satisfy the following properties: (1) The $E_i'C$ are mutually orthogonal, and (2) E_i' restricted to C is proportional to a restriction to C of a unitary operator.

To relate this characterization to detectability, note that the two properties imply that $(E_i')\, E_j'C$ is orthogonal to C if $i \neq j$ and $(E_i')^\dagger E_i'$ restricted to C is proportional to the identity on C. In other words, the $(E_i')^\dagger E_j'$ are detectable. This detectability condition applied to the original error set constitutes a second characterization of correctability, as given in Theorem 7.

Theorem 7. $\mathcal{E}$ is correctable if and only if the operators in the set $\mathcal{E}^\dagger\mathcal{E} = \{E_1^\dagger E_2 : E_i \in \mathcal{E}\}$ are detectable.

Before explaining the characterizations of correctability, we consider the situation of n qubits, where the characterization by detectability (Theorem 7) leads to a useful relationship between minimum distance and correctability of low-weight errors.

Theorem 8. If a code on n qubits has a minimum distance of at least $2e + 1$, then the set of errors of weight at most e is correctable.

This theorem follows by observing that the weight of $E_1^\dagger E_2$ is at most the sum of the weights of the E_i. As a result of this observation, the problem of finding good ways to correct all errors up to a maximum weight reduces to that of constructing codes with sufficiently high minimum distance. Thus, questions such as "what is the maximum dimension of a code of minimum distance d on n qubits?" are of great interest. As in the case of classical coding theory, this problem appears to be very difficult in general. Answers are known for small n (Calderbank et al. 1998), and there are asymptotic bounds (Ashikhmin and Litsyn 1999). Of course, for achieving low error probabilities, it is not necessary to correct all errors of weight $\leq e$, just almost all such errors. For example, the concatenated codes used for fault-tolerant quantum computation achieve this goal (see "Fault-Tolerant Quantum Communication and Computation" later in this article).

For the remainder of this section, we explain the characterizations of correctability. Using the conditions for detectability from the previous section, the condition for correctability in Theorem 7 is equivalent to

$$PE_i^\dagger E_j P = \lambda_{ij} P \ . \tag{39}$$

This condition is preserved under a linear change of basis for $\mathcal{E}$. That is, if A is any invertible matrix with coefficients a_{ij}, we can define new error operators $D_k = \sum_i E_i a_{ik}$. For the D_k, the left side of Equation (39) is

$$\begin{aligned} PD_k^\dagger D_l P &= P\left(\sum_{ij} \bar{a}_{ik} E_i^\dagger E_j a_{jl}\right)P \\ &= \sum_{ij} \bar{a}_{ik} a_{jl} \lambda_{ij} P \\ &= \left(A^\dagger \Lambda A\right)_{kl} P \ \ , \end{aligned} \tag{40}$$

where Λ is the matrix formed from the λ_{ij}. Using the fact that Λ is a positive semidefinite matrix (that is, for all x, $x^\dagger \Lambda x \geq 0$, and $\Lambda^\dagger = \Lambda$), we can choose A such that $A^\dagger \Lambda A$ is of the form $\begin{pmatrix} \mathbb{1} & 0 \\ 0 & 0 \end{pmatrix}$. In this matrix, the upper left block is the identity operator for some dimension.

An important consequence of invariance under a change of basis of error operators is that the set of errors correctable by a particular code and decoding procedure is linearly closed. Thus, if E and D are corrected by the decoding procedure, then so is $\alpha E + \beta D$. This observation also follows from the linearity of quantum mechanically implementable operations.

We explain the condition for correctability by using the subsystems interpretation of decoding procedures. For simplicity, assume that $\mathbb{1} \in \mathcal{E}$. To show that correctability of $\mathcal{E}$ implies detectability of all $E \in \mathcal{E}^\dagger \mathcal{E}$, suppose that we have a decoding procedure that recovers the information encoded in C after any of the errors in the set $\mathcal{E}$ have occurred. Every physically realizable decoding procedure can be implemented by first adding ancilla quantum systems in a prepared pure state to form a total system labeled T, then applying a unitary map U to the state of T, and finally separating T into a pair of systems (syn, Q), where "syn" corresponds to the syndrome subsystem and Q is a quantum system with the same dimension as the code that carries the quantum information after decoding. Denote the state space of the physical system containing C as $\mathcal{H}$ and the state space of system X by $\mathcal{H}_{\mathrm{X}}$, where X is any one of the other systems. Let V be the unitary operator that encodes information by mapping $\mathcal{H}_{\mathrm{Q}}$ onto $C \subseteq \mathcal{H}$. We have the following relationships:

$$\mathcal{H}_{\mathrm{Q}} \overset{V}{\leftrightarrow} C \subseteq \mathcal{H} \subseteq \mathcal{H}_{\mathrm{T}} \overset{U}{\leftrightarrow} \mathcal{H}_{\mathrm{syn}} \otimes \mathcal{H}_{\mathrm{Q}} \ . \tag{41}$$

Here, we used bidirectional arrows to emphasize that the operators V and U can be inverted on their range and therefore identify the states in their domains with the states in their ranges. The inclusion $\mathcal{H} \subseteq \mathcal{H}_{\mathrm{T}}$ implicitly identifies $\mathcal{H}$ with the subspace determined by the prepared pure state on the ancillas. The last state space of Equation (41) is expressed as a tensor product, which is the state space of the combined system (syn, Q).

For states of $\mathcal{H}_{\mathrm{Q}}$, we will write $|\psi\rangle = |\psi\rangle_{\mathrm{Q}} \overset{V}{\leftrightarrow} |\psi\rangle_{\mathrm{L}} \in C$. Because $\mathbb{1}$ is a correctable error, it must be the case that $|\psi\rangle_{\mathrm{L}} \overset{U}{\leftrightarrow} |0\rangle_{\mathrm{syn}}|\psi\rangle \in \mathcal{H}_{\mathrm{syn}} \otimes \mathcal{H}_{\mathrm{Q}}$ for some state $|0\rangle_{\mathrm{syn}}$. To establish this fact, use the linearity of the maps. In general,

$$\begin{aligned} |\psi\rangle_{\mathrm{L}} &\to E_i|\psi\rangle_{\mathrm{L}} \\ &\overset{U}{\leftrightarrow} |i\rangle_{\mathrm{syn}}|\psi\rangle \quad . \end{aligned} \tag{42}$$

The $|i\rangle_{\mathrm{syn}}$ need not be normalized or orthogonal. Let F be the subspace spanned by the $|i\rangle_{\mathrm{syn}}$. Then U induces an identification of $F \otimes \mathcal{H}_{\mathrm{Q}}$ with a subspace $\overline{C} \subseteq \mathcal{H}$.
This is the desired subsystem identification. We can then see how the errors act in this identification.

$$\begin{tikzcd} |\psi\rangle_{\mathrm{L}} \arrow[r, leftrightarrow] \arrow[d] & |0\rangle_{\mathrm{syn}}|\psi\rangle \\ E_i|\psi\rangle_{\mathrm{L}} \arrow[r, leftrightarrow] & |i\rangle_{\mathrm{syn}}|\psi\rangle \ . \end{tikzcd} \tag{43}$$

This means that for all $|\psi\rangle$ and $|\phi\rangle$,

$${}^{\mathrm{L}}\langle\psi|E_j^{\dagger}E_i|\phi\rangle_{\mathrm{L}} = {}^{\mathrm{syn}}\langle j|i\rangle_{\mathrm{syn}}\langle\psi|\phi\rangle \quad , \tag{44}$$

that is, all errors in $\mathcal{E}^{\dagger}\mathcal{E}$ are detectable.

Now, suppose that all errors in $\mathcal{E}^{\dagger}\mathcal{E}$ are detectable. To see that correctability of $\mathcal{E}$ follows, choose a basis for the errors so that $\lambda_{ij} = \delta_{ij}\lambda_i$ with $\lambda_i = 1$ for $i < s$ and $\lambda_i = 0$ otherwise. Define a subsystem identification by

$$|i\rangle_{\mathrm{sys}}|\psi\rangle \underset{\rightarrow}{W} E_i|\psi\rangle_{\mathrm{L}} \quad , \tag{45}$$

for $0 \le i < s$. By assumption and construction, ${}^{\mathrm{L}}\langle\psi|E_j^{\dagger}E_i|\psi\rangle_{\mathrm{L}} = \delta_{ij}$, which implies that W is unitary (after linear extension), and so this is a proper identification. For $i \ge s$, $E_i\,|\psi\rangle_{\mathrm{L}} = 0$, which implies that for states in the code, these errors have probability 0. Therefore, the identification can be used to successfully correct $\mathcal{E}$.

Constructing Codes

Stabilizer Codes. Most useful quantum codes are based on stabilizer constructions (Gottesman 1996, Calderbank et al. 1997). Stabilizer codes are useful because they make it easy to determine which Pauli-product errors are detectable and because they can be interpreted as special types of classical, linear codes. The latter feature makes it possible to use well-established techniques from the theory of classical error-correcting codes to construct good quantum codes.

A stabilizer code of length n for k-qubits (abbreviated as an $[[n, k]]$ code), is a 2^k-dimensional subspace of the state space of n-qubits that is characterized by the set of products of Pauli operators that leave each state in the code invariant. Such Pauli operators are said to stabilize the code. A simple example of a stabilizer code is the quantum repetition code introduced earlier. The code's states $\alpha|000\rangle + \beta|111\rangle$ are exactly the states that are unchanged after applying $\sigma_z^{(1)}\,\sigma_z^{(2)}$ or $\sigma_z^{(1)}\,\sigma_z^{(3)}$. To simplify the notation, we write $I = \mathbb{1}$, $X = \sigma_x$, $Y = \sigma_y$, and $Z = \sigma_z$. A product of Pauli operators can then be written as $ZIXI = \sigma_z^{(1)}\,\sigma_x^{(3)}$ (as an example of length 4) with the ordering determining which qubit is being acted upon by the operators in the product.

We can understand the properties of stabilizer codes by working out the example of the quantum repetition code with the stabilizer formalism. A stabilizer of the code is $S = \{ZZI, ZIZ\}$. Let $\overline{S}$ be the set of Pauli products that are expressible up to a phase as products of elements of S. For the repetition code, $\overline{S} = \{III, ZZI, ZIZ, IZZ\}$. $\overline{S}$ consists of all Pauli products that stabilize the code. The crucial property of S is that its operators commute, that is, for $A, B \in S$, $AB = BA$. According to results from linear algebra, it follows that the state space $\mathcal{H}$ can be decomposed into orthogonal subspaces $\mathcal{H}_\lambda$ such that for $A \in S$ and $|\psi\rangle \in \mathcal{H}_\lambda$, $A|\psi\rangle = \lambda(A)|\psi\rangle$. The $\mathcal{H}_\lambda$ are the common eigenspaces of S. The stabilizer code C defined by S is the subspace stabilized by the operators in S, which means that it is given by $\mathcal{H}_\lambda$ with $\lambda(A) = 1$. The subspaces for other $\lambda(A)$ have equivalent properties and are often included in the set of stabilizer codes. For the repetition code, the stabilized subspace is spanned by the logical basis $|000\rangle$ and $|111\rangle$. From the point of view of stabilizers, there are two ways in which a Pauli product B can be detectable: (1) if $B \in \overline{S}$ because, in this case, B acts as the identity on the code and (2) if B anticommutes with at least one member (say A) of S. To see that this statement is correct, let $|\psi\rangle$ be in the code. Then $A(B|\psi\rangle) = (AB)|\psi\rangle = -(BA)|\psi\rangle = -B(A|\psi\rangle) = -B|\psi\rangle$. Thus, $B|\psi\rangle$ belongs to $\mathcal{H}_\lambda$ with $\lambda(A) = -1$. Because this subspace is orthogonal to $C = \mathcal{H}_1$, B is detectable. We define the set of Pauli products that commute with all members of S as $\overline{S}^\perp$. Thus, B is detectable if either $B \notin \overline{S}^\perp$ or $B \in \overline{S}$. Note that because $\overline{S}$ consists of commuting operators, $\overline{S} \subseteq \overline{S}^\perp$.

To construct a stabilizer code that can correct all errors of weight at most one (a quantum one-error-correcting code), it suffices to find S with the minimum weight of nonidentity members of $\overline{S}^\perp$ being at least three ($3 = 2 \cdot 1 + 1$)—also refer to Theorem 8. In this case, we say that $\overline{S}^\perp$ has minimum distance 3. As an example, we can exhibit a stabilizer for the famous length-five one-error-correcting code for one qubit (Bennett et al. 1996, Laflamme et al. 1996):

$$S = \{X\,Z\,Z\,X\,I,\; I\,X\,Z\,Z\,X,\; X\,I\,X\,Z\,Z,\; Z\,X\,I\,X\,Z\} \quad . \tag{46}$$

As a general rule, it is desirable to exhibit the stabilizer minimally, which means that no member is the product up to a phase of some of the other members. In this case, the number of qubits encoded is $n - |S|$, where n is the length of the code and $|S|$ is the number of elements of S.

To obtain the correspondence between stabilizer codes and classical binary codes, we replace the symbols I, X, Y, and Z in a Pauli product by 00, 01, 10, and 11, respectively. Thus, the members of the stabilizer can be thought of as binary vectors of length $2n$. We use arithmetic modulo 2 for sums, inner products, and application of a binary matrix. Because the numbers modulo 2 ($\mathbb{Z}_2$) form a mathematical field, the basic properties of vector spaces and linear algebra apply to binary vectors and matrices. Thus, the stabilizer is minimal in the sense introduced above if the corresponding binary vectors are independent over $\mathbb{Z}_2$. Given two binary (column) vectors x and y of length 2 associated with Pauli products, the property of anticommuting is equivalent to$x^TBy = 1$, where B is the block diagonal $2n \times 2n$ matrix with 2×2 blocks given by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

This means that $\bar{S}^{\perp}$ can be identified with the set of vectors x such that $x^T By = 0$ for all binary vectors y associated with the members of S. It turns out that the inner product $\langle x, y\rangle = x^T By$ arises in the study of classical codes over the four-element mathematical field $GF(4)$, which can be represented by the vectors 00, 01, 10, and 11 with addition modulo 2 and a new multiplication operation. This relationship leads to the construction of many good stabilizer codes (Calderbank et al. 1998).

Conserved Quantities, Symmetries, and Noiseless Subsystems. Even though a physical system may be exposed to error, some of its properties are often not affected by the errors. If these conserved quantities can be identified with the defining quantities of qubits or other information units, error-free storage of information can be ensured without active intervention. This is the idea behind noiseless subsystems.

When do noiseless subsystems exist and how can they be constructed? The examples discussed in the previous sections show that a noiseless subsystem may be a subset of physical qubits, as in the trivial two-qubit example, or it may require a more abstract subsystem identification, as in the example of the three spin-1/2 particles. As will be explained, in both cases, there are quantities conserved by the errors that can be used to identify the noiseless subsystem.

A simple classical example for the use of conserved quantities consists of two physical bits subject to errors that either flip both bits or leave them alone. A quantity invariant under this noise model is the parity $P(s)$ of a state s of the two bits. The parity $P(s)$ is defined as the number of 1s in the bit string s reduced modulo 2: $P(00) = P(11) = 0$, and $P(01) = P(10) = 1$. Flipping both bits does not change the value of P. Consequently, the two values of P can be used to identify the two states of a noiseless bit. The syndrome subsystem can be associated with the value (nonconserved) of the first physical bit using the function defined by $F(0b) = 0$, $F(1b) = 1$. The corresponding subsystem identification is obtained by using the values of P and F as the states of the syndrome (left) and the noiseless information-carrying subsystem (right) according to $ab \leftrightarrow F(ab) \cdot P(ab)$.

In quantum systems, conserved quantities are associated with the presence of symmetries, that is, with operators that commute with all possible errors. In the trivial two-qubit example, operators acting only on qubit 2 commute with the error operators. In particular, if E is any one of the errors, $E\sigma_u^{(2)} = \sigma_u^{(2)}E$ for $u = x, y, z$. It follows that the expectations of $\sigma_u^{(2)}$ are conserved. That is, if ρ is the initial state (density matrix) of the two physical qubits and ρ' is the state after the errors acted, then $\mathrm{tr}\,\sigma_u^{(2)}\rho' = \mathrm{tr}\,\sigma_u^{(2)}\rho$. Because the state of qubit 2 is completely characterized by these expectations, it follows immediately that it is unaffected by the noise.

The trivial two-qubit example suggests a general strategy for finding a noiseless qubit: First, determine the commutant of the errors, which is the set of operators that commute with all errors. Then, find a subset of the commutant that is algebraically equivalent to the operators characterizing a qubit. The equivalence can be formulated as a one-to-one map f from qubit operators to operators in the commutant. For the range of f to be algebraically equivalent, f must be linear and satisfy $f(A^\dagger) = f(A)^\dagger$ and $f(AB) = f(A)f(B)$. Once such an equivalence is found, a fundamental theorem from the representation theory of finite dimensional operator algebras implies that a subsystem identification for a noiseless qubit exists (Knill et al. 2000, Viola et al. 2001).

The strategy can be applied to the example of three spin-1/2 particles subject to collective errors. One can determine the commutant by using the physical properties of spin to find the conserved quantities associated with operators in the commutant, as suggested in the box "Creating a Noiseless Subsystem from Three Spin-1/2 Particles" on page 205. Alternatively, observe that, by definition, this error model is symmetric under permutations of the particles. Therefore, the actions of these permutations on the state space form a group Π of unitary operators commuting with the errors. It is a fact that the commutant of the set of collective errors consists of the linear combinations of operators in Π. With respect to the group Π, one can immediately determine the space $V_{3/2}$ of symmetric states, that is, those that are invariant under the permutations. It is spanned by

$$\left|\uparrow\uparrow\uparrow\right\rangle,\ \frac{1}{\sqrt{3}}\left(\left|\uparrow\uparrow\downarrow\right\rangle+\left|\uparrow\downarrow\uparrow\right\rangle+\left|\downarrow\uparrow\uparrow\right\rangle\right),\ \frac{1}{\sqrt{3}}\left(\left|\uparrow\downarrow\downarrow\right\rangle+\left|\downarrow\uparrow\downarrow\right\rangle+\left|\downarrow\downarrow\uparrow\right\rangle\right),\ \left|\downarrow\downarrow\downarrow\right\rangle\ . \tag{47}$$

A basic result from the representation theory of groups implies that the projection onto $V_{3/2}$ is given by $P_{3/2} = (1/6)\Sigma_{g\in\Pi}g$. The orthogonal complement $V_{1/2}$ of $V_{3/2}$ is invariant under Π and can be analyzed separately. With the subsystem identification of Equation (26) already in hand, one can see that the permutation π_1, which permutes the spins according to $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$, acts on the noiseless qubit, by applying $Z_{240^\circ} = e^{-i\sigma_z 2\pi/3}$, a 240° rotation around the z-axis. Similarly, the permutation π_2, which exchanges the last two spins, acts as σ_x on the qubit. To make them algebraically equivalent to the corresponding qubit operators, it is necessary to eliminate their action on $V_{3/2}$ by projecting onto $V_{1/2}$: $\pi_1' = (1 - P_{3/2})\pi_1$ and $\pi_2' = (1 - P_{3/2})\pi_2$. Sums of products of π_1' and π_2' are equivalent to the corresponding sums of products of Z_{240° and σ_x, which generate all qubit operators. To get the subsystem identification of Equation (26), one can start with a common eigenstate $|\psi\rangle$ of π_1' (a z-rotation on the noiseless qubit) and $2J_z$ (the syndrome subsystem's σ_z) with eigenvalues $e^{-i2\pi/3}$ and 1, respectively. The choice of eigenvalues implies that $|\psi\rangle \leftrightarrow |\uparrow\rangle \cdot |\circ\rangle$ in the desired identification. We can obtain the other logical states of the syndrome spin 1/2 and the noiseless qubit by applying π_2', $2J_x$, and $\pi_2' 2J_x$ to $|\psi\rangle$, which act by flipping the states of the qubit or the syndrome spin. This method for obtaining the subsystem identification generalizes to other operator equivalences and error operators.

Fault-Tolerant Quantum Communication and Computation

The utility of information and information processing depends on the ability to implement large numbers of information units and information-processing operations. We say that an implementation of information processing is scalable if the implementation can realize arbitrarily many information units and operations without loss of accuracy and with physical resource overheads that are polynomial (or efficient) in the number of information units and operations. Scalable information processing is achieved by implementing information fault-tolerantly.

One of the most important results of the work in quantum error-correction and fault-tolerant computation is the accuracy threshold theorem, according to which scalability is possible, in principle, for quantum information.

Theorem 9. Assume the requirements for scalable QIP (see below). If the error per gate is less than a threshold, then it is possible to efficiently quantum-compute to arbitrary accuracy.

Requirements for Scalable QIP. The value of the threshold accuracy (or error) depends strongly on which set of requirements is used—in particular, the error model that is assumed. The requirements are closely related to the basic requirements for constructing a quantum information processor (DiVincenzo 2000) but have to include explicit assumptions on the error model and on the temporal and spatial aspects of the available quantum control:

Scalable physical systems. It is necessary to access physical systems that are able to support qubits or other basic units of quantum information. The systems must be scalable; that is, they must be able to support any number of independent qubits.

State preparation. One must be able to prepare any qubit in the standard initial state $|0\rangle$. Any preexisting content is assumed to be lost, as would happen if, for example, the qubit is first discarded and then replaced by a prepared one. The condition can be weakened; that is, it is sufficient that a large fraction of the qubits can be prepared in this way.

Measurement. Being able to measure any qubit in the logical basis is a requirement. Again, it is sufficient that a large enough fraction of the qubits are measurable. For solving computational problems with deterministic answers, the standard projective measurement can be replaced by weak measurements that return a noisy number whose expectation is the probability that a qubit is in the state $|1\rangle$ (Laflamme et al. 2001).

Quantum control. One must be able to implement a universal set of unitary quantum gates acting on a small number (usually, at most, two at a time) of qubits. For most accuracy thresholds, it is necessary to be able to apply the quantum control in parallel to any number of disjoint pairs of qubits. This parallelism requirement can be weakened if a nearly noiseless quantum memory is available. The requirement that it be possible to apply two-qubit gates to any pair of qubits is unrealistic given the constraints of three-dimensional space. Work on how to deal with this problem is ongoing (Aharonov and Ben-Or 1999). The universality assumption can be substantially weakened by replacement of some or all unitary quantum gates with operations to prepare special states or by additional measurement capabilities. See, for example, Michael Nielsen (2001) and the references therein.

Errors. The error probability per gate must be below a threshold and satisfy independence and locality properties (refer to the section "Error Models"). The definition of gate includes the "no-op," which is the identity operation implemented over the time required for a computational step. For the most pessimistic, independent, local error models, the error threshold is above $\sim 10^{-6}$; for the independent depolarizing errors, it is believed to be better than 10^{-4} (Gottesman and Preskill 1999). For some special error models, the threshold is substantially higher. For example, for the independent "erasure" error model, where error events are always detected, the threshold is above .01, and for an error model whose errors are specific, unintentional measurements in the standard basis of a qubit, the threshold is 1 (Knill et al. 2000). The threshold is also well above .01 when the goal is only to transmit quantum information through noisy quantum channels (Briegel et al. 1998).

Realizing Fault Tolerance. The existing proofs of the accuracy threshold theorems consist of explicit instructions for building a scalable quantum information processor and analyses of its robustness against the assumed error model. The instructions for realizing scalable computation are based on the following simple idea. Suppose that the error rate per operation for some way of realizing qubits is p. We can use these qubits and a quantum error-correcting code to encode logical qubits for which the storage error rate is reduced. For example, if a one-error correcting code is used, the error rate per storage interval for the logical qubits is expected to be $\leq cp^2$ for some constant c. Suppose that we can show how to implement encoded operations, preparations, measurement, and the subroutines required for error correction such that this inequality is now valid for each basic encoded step, perhaps for a larger constant C. Suppose furthermore that the errors for the encoded information still satisfy the assumed error model. The newly defined logical qubits then have an error rate of $\leq Cp^2$, which is less than p

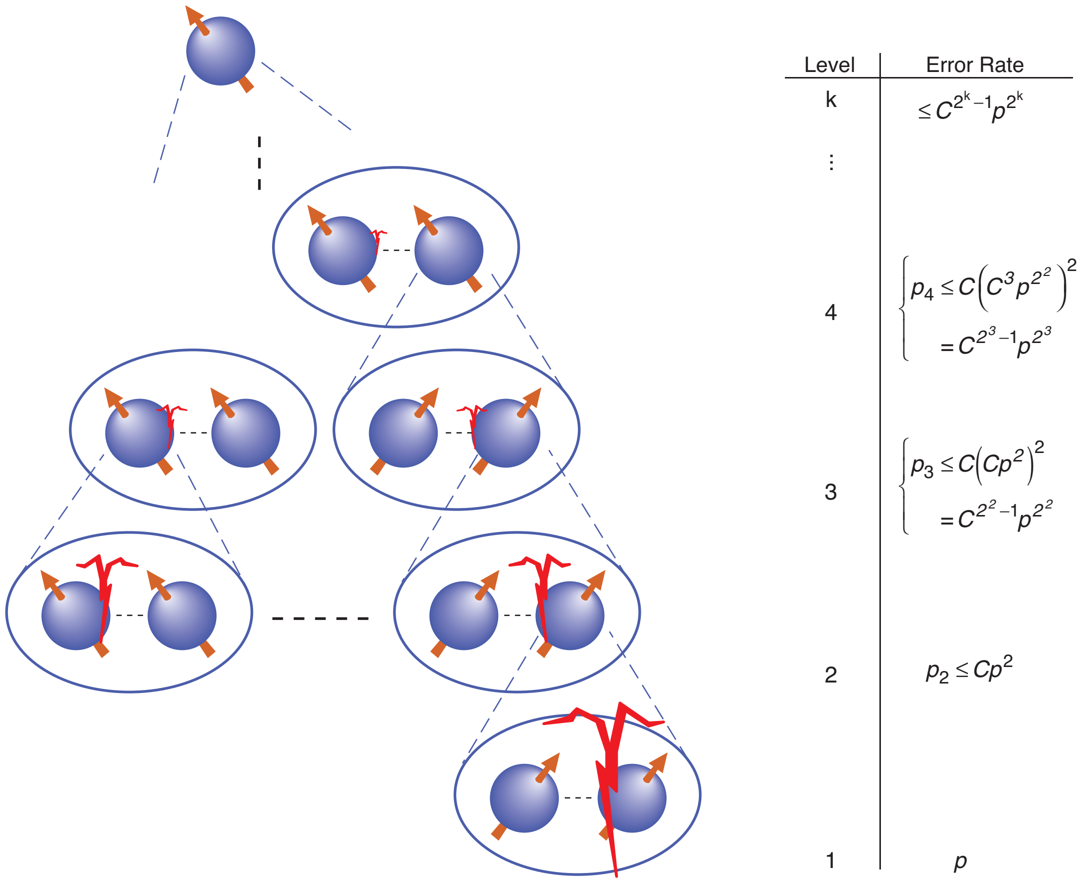

Level
Error Rate
k
$\leq C^{2^k-1}p^{2^k}$
⋮
4
$\begin{cases} p_4 \leq C\left(C^3p^{2^2}\right)^2 \\ = C^{2^3-1}p^{2^3} \end{cases}$
3
$\begin{cases} p_3 \leq C\left(Cp^2\right)^2 \\ = C^{2^2-1}p^{2^2} \end{cases}$
2
$p_2 \leq Cp^2$
1
p

Figure 8. Schematic Representation of Concatenation

The bottom level represents qubits realized more or less directly in a physical system. Each next level represents logical qubits defined by means of subsystems in terms of the previous level's qubits. More efficient subsystems might represent multiple qubits in one code block rather than the one qubit per code block shown here.

for $p < 1/C$. We can use the newly realized qubits as a foundation for making higher-level logical qubits. The result is multiple levels of encodings. In the next level (level 2), the error rate is $\leq C^3p^4$, and after k iterations, it is $\leq C^{2^k-1}p^{2^k}$, a doubly exponentially decreasing function of k. This procedure is called concatenation (refer to Figure 8). Because the complexity, particularly the number of physical qubits needed for each final logical qubit, grows only singly exponentially in k, the procedure is efficient. Specifically, to achieve a logical error of ε per operation requires of the order of $|\log(\varepsilon)|^r$ resources per logical qubit for some finite r. In practice, this simple idea is still dauntingly complex, but there is hope that, for realistic errors in physical systems and by cleverly trading off different variations of these techniques, much of the theoretical complexity can be avoided (Steane 1999).

Many important developments and ideas of quantum information were ultimately needed to realize encoded operations, preparations, measurements, and error correction subroutines that behave well with respect to concatenation. Stabilizer codes provide a particularly nice setting for implementing many of these techniques. One reason is that good stabilizer codes are readily constructed. Another is that they enable encoding operations in a way that avoids spreading errors between the qubits of a single code word (Gottesman 1998). In addition, there are many tricks based on teleportation that can be used to maintain the syndrome subsystems in acceptably low error states and to implement general operations systematically (Gottesman and Chuang 1999). To learn more about all these techniques, see the textbook by Nielsen and Isaac Chuang (2001) and the works of Daniel Gottesman (1998) and John Preskill (1998).

Concluding Remarks

The advancements in quantum error-correction and fault-tolerant QIP have shown that, in principle, scalable quantum computation is achievable. This is a crucial result because it suggests that experimental efforts in QIP will eventually lead to more than a few small-scale applications of quantum information to communication and problems with few qubits. However, the general techniques for achieving scalability that are known are difficult to realize. Existing technologies are far from achieving sufficient accuracy even for just two qubits—at least in terms of the demands of the usual accuracy-threshold theorems. There is hope that more optimistic thresholds can be shown to apply if one takes into consideration the specific constraints of a physical device, better understands the dominant sources of errors, and exploits tailor-made ways of embedding quantum information into subsystems. Current work in this area is focused on finding such methods of quantum error control. These methods include approaches to error control not covered in this article—for example, techniques for actively turning off the error-inducing environmental interactions (Viola and Lloyd 1998, Viola et al. 1999) and modifications to controlling quantum systems that eliminate systematic and calibration errors (Levitt 1982, Cummins and Jones 1999). Further work is also needed to improve the thresholds for the more pessimistic error models and for developing more-efficient scalability schemes. ■

Contact Information

E. Knill: knill@lanl.gov

R. Laflamme: laflamme@iqc. ca

A. Ashikhmin:
aea@research.bell-labs.com

H. Barnum: barnum@lanl.gov

L. Viola: viola@lanl.gov

W. H. Zurek: whz@lanl.gov

Further Reading

Aharonov, D., and M. Ben-Or. 1996. Fault-Tolerant Quantum Computation with Constant Error Rate. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computation* (*STOC*). New York: ACM Press.

———. 1999. Fault-Tolerant Quantum Computation with Constant Error. [Online]: http://eprints.lanl.gov. (quant-ph/9906129).

Ashikhmin, A., and S. Litsyn. 1999. Upper Bounds on the Size of Quantum Codes. *IEEE Trans. Inf. Theory* **45**: 1206.

Bennett, C. H., D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. 1996. Mixed State Entanglement and Quantum Error-Correcting Codes. *Phys. Rev. A* **54**: 3824.

Braunstein, S. L., and H.-K. Lo, eds. 2000. Special Focus Issue on Experimental Proposals for Quantum Computation: Forward. *Fortschr. Phys*. **48** (9–11): 76

Briegel, H.-J., W. Dür, J. I. Cirac, and P. Zoller. 1998. Quantum Repeaters for Communication. [Online]: http://eprints.lanl.gov. (quant-ph/9803056).

Calderbank, A. R., E. M. Rains, P. W. Shor, and N. J. A. Sloane. 1997. Quantum Error Correction and Orthogonal Geometry. *Phys. Rev. A* **78**: 405.

———. 1998. Quantum Error Correction via Codes over gf(4). *IEEE Trans. Inf. Theory* **44**: 1369.

Cummins, H. K., and J. A. Jones. 2000. Use of Composite Rotations to Correct Systematic Errors in NMR Quantum Computation. *New J. Phys*. **2**: 6.

DiVincenzo, D. P. 2000. The Physical Implementation of Quantum Computation. *Fortschr. Phys*. **48**: 771.

Gottesman, D. 1996. A Class of Quantum Error-Correcting Codes Saturating the Quantum Hamming Bound. *Phys. Rev. A* **54**: 1862.

———. 1998. A Theory of Fault-Tolerant Quantum Computation. *Phys. Rev. A* **57**: 127.

Gottesman, D., and I. L. Chuang. 1999. Demonstrating the Viability of Universal Quantum Computation Using Teleportation and Single-Qubit Operations. *Nature* **402**: 390.

Kitaev, A. Yu. 1997. Quantum Error Correction with Imperfect Gates. In *Quantum Communication and Computing and Measurement*. Edited by 0. Hirota et al. New York: Plenum.

Knill, E., and R. Laflamme. 1996. “Concatenated Quantum Codes.” [Online]: http://eprints.lanl.gov. (quant-ph/9608012).

Knill, E., R. Laflamme, and G. Milburn. 2000. “Thresholds for Linear Optics Quantum Computation.” [Online]: http://eprints.lanl.gov. (quant-ph/0006120).

Knill, E., R. Laflamme, and W. H. Zurek. 1998a. Resilient Quantum Computation. *Science* **279**: 342.

———. 1998b. Resilient Quantum Computation: Error Models and Thresholds. *Proc. R. Soc. London, Ser. A* **454**: 365.

Knill, E., R. Laflamme, and L. Viola. 2000. Theory of Quantum Error Correction for General Noise. *Phys. Rev. Lett.* **84**: 2525.

Kraus, K. 1983. States, Effects and Operations: Fundamental Notions of Quantum Theory. *Lecture Notes in Physics*. Vol. 190. Berlin: Springer-Verlag.

Laflamme, R., C. Miquel, J-P. Paz, and W. H. Zurek. 1996. Perfect Quantum Error-Correcting Code. *Phys. Rev. Lett.* **77**: 198.

Levitt, M. H. 1982. Symmetrical Composite Pulse Sequences for NMR Population-Inversion. I. Compensation for Radiofrequency Field Inhomogeneity. *J Mag. Res*. **48**:234.

Nielsen, M. A. 2001. Universal Quantum Computation Using Only Projective Measurement, Quantum Memory, and Preparation of the $|0\rangle$ State. [Online]: http://eprints.lanl.gov. (quant-ph/0108020).

Nielsen, M. A., and I. L. Chuang. 2001. *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge University Press.

Preskill, J. 1998. Reliable Quantum Computers. *Proc. R. Soc. Lond*on*, Ser. A* **454**: 385.

Schumacher, B. 1996. Sending Entanglement through Noisy Quantum Channels. *Phys. Rev. A* **54**: 2614.

Shor, P. W. 1995. Scheme for Reducing Decoherence in Quantum Computer Memory. *Phys. Rev. A* **2**: 2493.

———. 1996. Fault-Tolerant Quantum Computation. In *Proceedings of the 37th Symposium on the Foundations of Computer Science (FOCS)*. p. 56. Los Alamitos, CA: IEEE Press.

Steane, A., 1996. Multiple Particle Interference and Quantum Error Correction. *Proc. R. Soc. London, Ser. A* **452**: 2551.

———. 1999. Efficient Fault-Tolerant Quantum Computing. *Nature* **399**: 124.

Viola, L., and S. Lloyd. Dynamical Suppression of Decoherence in Two-State Quantum Systems. *Phys. Rev. A* **58**: 2733.

Viola, L., E. Knill, R. Laflamme. 2001. Constructing Qubits in Physical Systems. *J. Phys. A* **34** (35): 7067.

Viola, L., E. Knill, and S. Lloyd. 1999. Dynamical Decoupling of Open Quantum Systems. *Phys. Rev. Lett*. **82**: 2417.

Glossary

Bit. The basic unit of deterministic information. It is a system that can be in one of two possible states, 0 and 1.

Bit string. A sequence of 0s and 1s that represents a state of a sequence of bits. The bit strings are words in the binary alphabet.

Classical information. The type of information based on bits and bit strings and, more generally, on words formed from finite alphabets. This is the information used for communication between people. Classical information can refer to deterministic or probabilistic information, depending on the context.

Code. A set of states that can be used to represent information. The set of states needs to have the properties of the type of information to be represented. The code is usually a subset of the states of a given system Q. It is then a Q-code or a code on Q. If information is represented by a state in the code, Q is said to carry the information.

Code word. A state in a code. The term is primarily used for classical codes defined on bits or systems with nonbinary alphabets.

Concatenation. An iterative procedure in which higher-level logical information units are implemented in terms of lower-level units.

Control error. An error due to nonideal control in applying operations or gates.

Communication channel. A means for transmitting information from one place to another. It can be associated with a physical system in which the information to be transmitted is stored by the sender. The system is subsequently conveyed to the receiver, who can then make use of the information.

Correctable error set. For a given code, a set of errors such that there is an implementable procedure R that, after any one of the errors E acts on a state x in the code, returns the system to the state $x = REx$. What procedures are implementable depends on the type of information represented by the system and, if it is a physical system, its physics.

Decoding. The process of transferring information from an encoded form to its "natural" form. In the context of error correction, decoding is often thought of as consisting of two steps: one which removes the errors' effects (sometimes called the recovery procedure) and one that extracts the information (often also called decoding in a narrower sense).

Depolarizing errors. An error model for qubits in which random Pauli operators are applied independently to each qubit.

Detectable error. For a given code, an error that has no effect if the state is observed to have remained in the code. If the state is no longer in the code, the error is said to have been detected, and the state no longer represents valid information.

Deterministic information. The type of information based on bits and bit strings. This is the same as classical information but explicitly excludes probabilistic information.

Encoding. The process of transferring information from its natural form to an encoded form. It requires an identification of the valid states associated with the information and the states of a code. The process acts on an information unit and replaces it with the system whose state space contains the code.

Environment. In the context of information encoded in a physical system, it refers to other physical systems that may interact with the information-carrying system.

Environmental noise. Noise due to unwanted interactions with the environment.

Error. Any unintended effect on the state of a system, particularly in storing or otherwise processing information.

Error basis. A set of state transformations that can be used to represent any error. For quantum systems, errors can be represented as operators acting on the system's state space, and an error basis is a maximal, linearly independent set of such operators.

Error control. The term for general procedures that limit the effects of errors on information represented in noisy, physical systems.

Error correction. The process of removing the effects of errors on encoded information.

Error-correcting code. A code with additional properties that enable a decoding procedure to remove the effects of the dominant sources of errors on encoded information. Any code is error correcting for some error model in this sense. To call a code error correcting emphasizes the fact that it was designed for this purpose.

Error model. An explicit description of how and when errors happen in a given system. Typically, a model is specified as a probability distribution over error operators. More general models may need to be considered, particularly in the context of fault-tolerant computation, for which correlations in time are important.

Fault tolerance. A property of encoded information that is being processed with gates. It means that errors occurring during processing, including control errors and environmental noise, do not seriously affect the information of interest.

Gate. An operation applied to information for the purpose of information processing.

Hamming distance. The Hamming distance between two binary words (sequences of 0 and 1) is the number of positions in which the two words disagree.

Hilbert space. An n-dimensional Hilbert space consists of all complex n-dimensional vectors. A defining operation in a Hilbert space is the inner product. If the vectors are thought of as column vectors, then the inner product $\langle x, y\rangle$ of x and y is obtained by forming the conjugate transpose $x^\dagger$ of x and calculating $\langle x, y\rangle = x^\dagger y$. The inner product induces the usual norm $|x|^2 = \langle x, x\rangle$.

Information. Something that can be recorded, communicated, and computed with. Information is fungible, which implies that its meaning can be identified regardless of the particulars of the physical realization. Thus, information in one realization (such as ink on a sheet of paper) can be easily transferred to another (for example, spoken words). Types of information include deterministic, probabilistic, and quantum information. Each type is characterized by information units, which are abstract systems whose states represent the simplest information of this type. These define the natural representation of the information. For deterministic information, the unit is the bit, whose states are symbolized by 0 and 1. Information units can be put together to form larger systems and can be processed with basic operations acting on a small number of units at a time.

Length. For codes on n basic information units, the length of the code is n.

Minimum distance. The smallest number of errors that is not detectable by a code. In this context, the error model consists of a set of error operators without specified probabilities. Typically, the concept is used for codes on n information units, and the error model consists of operators acting on any one of the units. For a classical binary code, the minimum distance is the smallest Hamming distance between two code words.

Noise. Any unintended effect on the state of a system, particularly an effect with a stochastic component due to incomplete isolation of the system from its environment.

Operator. A function transforming the states of a system. Operators may be restricted, depending on the system's properties. For example, operators acting on quantum systems are always assumed to be linear.

Pauli operators. The Hermitian matrices σ_x, σ_y, and σ_z—refer to Equation (7)—acting on qubits. It is often convenient to consider the identity operator to be included in the set of Pauli operators.

Physical system. A system explicitly associated with a physical device or particle. The term is used to distinguish between abstract systems used to define a type of information and specific realizations, which are subject to environmental noise and errors due to other imperfections.

Probabilistic bit. The basic unit of probabilistic information. It is a system whose state space consists of all probability distributions over the two states of a bit. The states can be thought of as describing the outcome of a biased coin flip before the coin is flipped.

Probabilistic information. The type of information obtained when the state spaces of deterministic information are extended with arbitrary probability distributions over the deterministic states. This is the main type of classical information with which quantum information is compared.

Quantum information. The type of information obtained when the state space of deterministic information is extended with arbitrary superpositions of deterministic states. Formally, each deterministic state is identified with one of an orthonormal basis vector in a Hilbert space, and superpositions are unit-length vectors that are expressible as complex linear sums of the chosen basis vectors. Ultimately, it is convenient to extend this state space again by permitting probability distributions over the quantum states. This is still called quantum information.

Qubit. The basic unit of quantum information. It is the quantum extension of the deterministic bit; that is, its state space consists of the unit-length vectors in a two-dimensional Hilbert space.

Repetition code. The classical, binary repetition code of length n consists of the two words 00 ... 0 and 11 ... 1. For quantum variants of this code, one applies the superposition principle to obtain the states consisting of all unit-length complex linear combinations of the two classical code words.

Scalability. A property of physical implementations of information processing that implies that there are no bounds on accurate information processing. That is, arbitrarily many information units can be realized, and they can be manipulated for an arbitrarily long amount of time without loss of accuracy. Furthermore, the realization is polynomially efficient in terms of the number of information units and gates used.

States. The set of states for a system characterizes the system's behavior and possible configurations.

Subspace. For a Hilbert space, a subspace is a linearly closed subset of the vector space. The term can be used more generally for a system Q of any information type: A subspace of Q or, more specifically, of the state space of Q is a subset of the state space that preserves the properties of the information type represented by Q.

Subsystem. A typical example of a subsystem is the first (qu)bit in a system consisting of two (qu)bits. In general, to obtain a subsystem of system Q, one first selects a subset C of Q's state space and then identifies C as the state space of a pair of systems. Each member of the pair is then a subsystem of Q. Restrictions apply, depending on the types of information carried by the system and subsystems. For example, if Q is quantum and so are the subsystems, then C has to be a linear subspace and the identification of the subsystems' state space with C has to be unitary.

Subsystem identification. The mapping or transformation that identifies the state space of two systems with a subset *C* of states of a system Q. In saying that L is a subsystem of Q, we also introduce a second subsystem and identify the state space of the combined system with the subset of states *C*.

Syndrome. One of the states of a syndrome subsystem. It is often used more narrowly for one of a distinguished set of basis states of a syndrome subsystem.

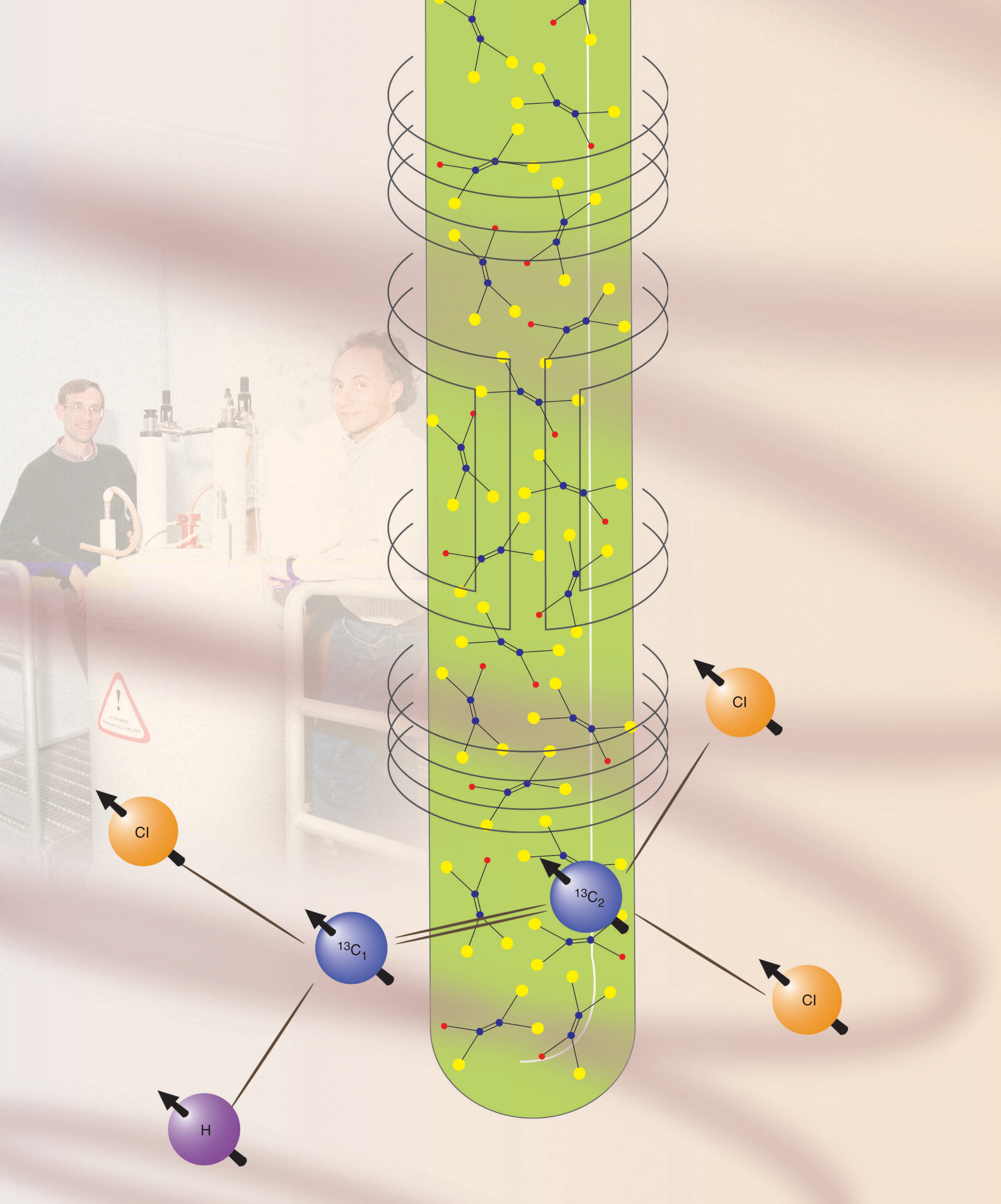
Syndrome subsystem. In identifying an information-carrying subsystem in the context of error correction, the other member of the pair of subsystems required for the subsystem identification is called the syndrome subsystem. The terminology comes from classical error correction, in which the syndrome is used to determine the most likely error that has occurred.

System. An entity that can be in any of a specified number of states. An example is a desktop computer whose states are determined by the contents of its various memories and disks. Another example is a qubit, which can be thought of as a particle whose state space is identified with complex, two-dimensional length-one vectors. Here, a system is always associated with a type of information, which in turn determines the properties of the state space**.** For example, for quantum information, the state space is a Hilbert space. For deterministic information, it is a finite set called an alphabet.

Twirling. A randomization method for ensuring that errors act like a depolarizing error model. For one qubit, it involves applying a random Pauli operator before the errors occur and then undoing the operator by applying its inverse.

Unitary operator. A linear operator U on a Hilbert space that preserves the inner product. That is, for all x and y, $\langle Ux, Uy\rangle = \langle x, y\rangle$. If U is given in matrix form, then this condition is equivalent to $U^{\dagger}U = \mathbb{1}$.

Weight. For a binary word, the weight is the number of 1s in the word. For an error operator acting on n systems by applying an operator to each one of them, the weight is the number of nonidentity operators applied.


Cl
Cl
$^{13}C_2$
$^{13}C_1$
Cl
H

NMR

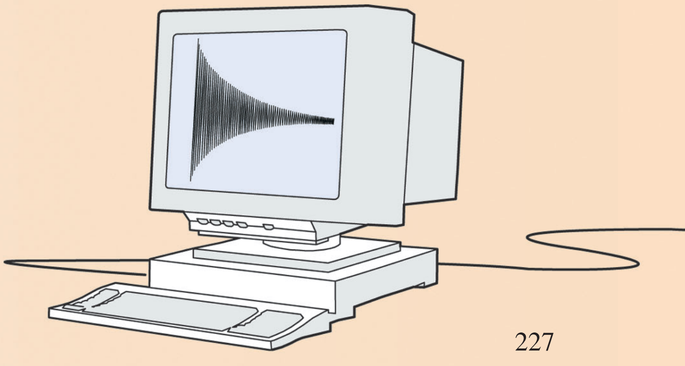
and Quantum Information Processing

Raymond Laflamme, Emanuel Knill, David G. Cory, Evan M. Fortunato, Timothy F. Havel,
Cesar Miquel, Rudy Martinez, Camille J. Negrevergne, Gerardo Ortiz, Marco A. Pravia,
Yehuda Sharf, Suddhasattwa Sinha, Rolando Somma, and Lorenza Viola

Using quantum physics to represent and manipulate information makes possible surprising improvements in the efficiency with which some problems can be solved. But can these improvements be realized experimentally? If we consider the history of implementing theoretical ideas about classical information and computation, we find that, initially, small numbers of simple devices were used to explore the advantages and difficulties of information processing. For example, in 1933, Atanasoff and his colleagues at the Iowa State College were able to implement digital calculations using about 300 vacuum tubes (Zalta 2002). Although the device was never practical because its error rate was too large, it was probably the first instance of a programmable computer using vacuum tubes, and it opened the way for more stable and reliable devices. Progress toward implementing quantum information processors is also initially confined to limited capacity and error-prone devices.

There are numerous proposals for implementing quantum information processing (QIP) prototypes. To date, however, only three of them have been used to successfully manipulate more than one qubit: cavity quantum electrodynamics (cavity QED), ion traps, and nuclear magnetic resonance (NMR) with molecules in a liquid (or liquid-state NMR). QIP devices are difficult to realize because of an intrinsic conflict between two of the most important requirements: On the one hand, it is necessary for the device to be well isolated from, and therefore interact only weakly with, its environment; otherwise, the crucial quantum correlations on which the advantages of QIP are based are destroyed. On the other hand, it is necessary for the different parts of the device to interact strongly with each other and for some of them to be coupled strongly with the measuring device, which is needed to read out “answers.” That few physical systems have these properties naturally is apparent from the absence of obvious quantum effects in the macroscopic world.

One system whose properties constitute a reasonable compromise between the two requirements consists of the nuclear spins in a molecule in the liquid state. The spins, particularly those with spin 1/2, provide a natural representation of quantum bits.



They interact weakly but reliably with each other, and the effects of the environment are often small enough. The spins can be controlled with radio-frequency (rf) pulses and observed with measurements of the magnetic fields they generate. Liquid-state NMR has so far been used to demonstrate control of up to seven physical qubits.

It is important to remember that the idea of QIP is less than two decades old, and, with the notable exception of quantum cryptography, experimental proposals and efforts aimed at realizing modern QIP began only in the last five years of the 20th century. Increasingly advanced experiments are being implemented. But from an information processing point of view, we are a long way from using quantum technology to solve an independently posed problem not solvable on a standard personal computer—a typical classical computer. In order to get to the point where such problems can be solved by QIP, current experimental efforts are devoted to understanding the behavior of and the methods for controlling various quantum systems, as well as ways of overcoming their limitations. The work on NMR QIP has focused on the control of quantum systems by algorithmically implementing quantum transformations as precisely as possible. Within the limitations of the device, this approach has been surprisingly successful—thanks to the many scientists and engineers who have perfected NMR spectrometers over the past 50 years.

After a general introduction to NMR, we give the basics of implementing quantum algorithms. We describe how qubits are realized and controlled with rf pulses, their internal interactions, and gradient fields. A peculiarity of NMR is that the internal interactions (given by the internal Hamiltonian) are always on. We discuss how they can be effectively turned off with the help of a standard NMR method called refocusing. Liquid-state NMR experiments are done at room temperature, leading to an extremely mixed (that is, nearly random) initial state. Despite this high degree of randomness, it is possible to investigate QIP because the relaxation time (the time scale over which useful signal from a computation is lost) is sufficiently long. We explain how this feature leads to the crucial ability of simulating a pure (nonrandom) state by using pseudopure states. We discuss how the answer provided by a computation is obtained by measurement and how this measurement differs from the ideal, projective measurement of QIP. We then give implementations of some simple quantum algorithms with a typical experimental result. We conclude with a discussion of what we have learned from NMR QIP so far and what the prospects are for future NMR QIP experiments. For an elementary, device-independent introduction to quantum information and definitions of the states and operators used here, see the article “Quantum Information Processing” on page 2 .

Liquid-State NMR

NMR Basics. Many atomic nuclei have a magnetic moment, which means that, like small bar magnets, they respond to and can be detected by their magnetic fields. Although single nuclei are impossible to detect directly by these means with currently available technology, if sufficiently many are available so that their contributions to the magnetic field add, they can be observed as an ensemble. In liquid-state NMR, the nuclei belong to atoms forming a molecule, a very large number of which are dissolved in a liquid. An example is carbon-13-labeled trichloroethylene (TCE)—see Figure 1. The hydrogen nucleus (that is, the proton) of each TCE molecule has a relatively strong magnetic moment. When the sample is placed in a powerful external magnetic field, each proton’s spin prefers to align itself with the field. It is possible to induce the spin direction to tip off-axis by means of rf pulses, at which point the effect of the static field is to induce a rapid precession of the proton spins. In this introduction, precession refers to a rotation of a spin direction around the main axis, here the z-axis, as determined by the external magnetic field. The precession frequency ω is often called the Larmor frequency and is linearly related to the strength B of the external field: $\omega = \mu B$, where μ is the magnetic moment. For the proton, the magnetic moment is 42.7 megahertz per tesla (MHz/T), so at a typical field of $B = 11.7$ tesla, the precession frequency is 500 megahertz. The magnetic field produced by the precessing protons induces oscillating currents in a coil judiciously placed around the sample and "tuned" to the precession frequency, allowing observation of the entire ensemble of protons by magnetic induction. This is the fundamental idea of NMR. The device that applies the static magnetic field and rf control pulses and that detects the magnetic induction is called an NMR spectrometer—see Figure 2.

Magnetic induction by nuclear spins was observed for the first time by Edward Purcell and coworkers (1946) and Felix Bloch (1946). This achievement opened a new field of research, leading to many important applications, such as molecular structure determination, dynamics studies both in the liquid and solid state (Ernst et al. 1994), and magnetic resonance imaging (Mansfield and Morris 1982). The application of NMR to QIP is related to methods for determining molecular structure by NMR. Many of the same techniques are used in QIP, but instead of using uncharacterized molecules, specific ones with well-defined nuclear spins are synthesized. In this setting, one can manipulate the nuclear spins as quantum information so that it becomes possible to experimentally demonstrate the fundamental ideas of QIP.

Perhaps the clearest example of early connections of NMR to information theory is the spin echo phenomenon (Hahn 1950). When the static magnetic field is not homogeneous (that is, it is not constant across the sample), the spins precess at different frequencies, depending on their location in the sample. As a result, the magnetic induction signal rapidly vanishes because the magnetic fields produced by the spins are no longer aligned and therefore do not add. The spin echo is used to refocus this effect by inverting the spins, an action that effectively reverses their precession until they are all aligned again. Based on spin echoes, the idea of using nuclear spins for (classical) information storage was suggested and patented by Arthur Anderson et al. (1955) and Anderson and Erwin Hahn (1955).

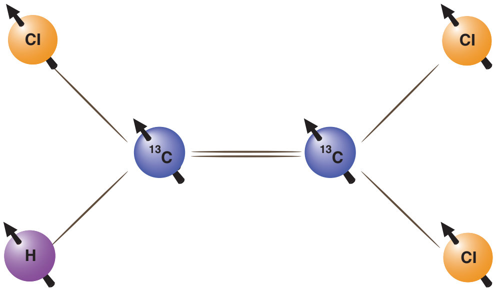


Figure 1. Schematic of a Typical Molecule (Trichloroethylene) Used for QIP

There are three useful nuclei for realizing qubits. They are the proton (H) and the two carbons (^{13}C). The molecule is "labeled," which means that the nuclei are carefully chosen isotopes. In this case, the normally predominant isotope of carbon, ^{12}C (a spin-zero nucleus), is replaced by ^{13}C, which has spin 1/2.

NMR spectroscopy would not be possible if it were not for relatively long "relaxation" times. Relaxation is the process that tends to realign the nuclear spins with the field and randomize their phases, an effect that leads to complete loss of the information represented in such a spin. In liquid state, relaxation times of the order of seconds are common and attributed to the weakness of nuclear interactions and a fast averaging effect associated with the rapid, tumbling motions of molecules in the liquid state.

Currently, off-the-shelf NMR spectrometers are robust and straightforward to use. The requisite control is to a large extent computerized, so most NMR experiments involve few custom adjustments after the sample has been obtained. Given that the underlying nature of the nuclear spins is intrinsically quantum mechanical, it is not surprising that, soon after Shor's discovery of the quantum factoring algorithm, NMR was studied as a potentially useful device for QIP.

A Brief Survey of NMR QIP. Concrete and workable proposals for using liquid-state NMR for quantum information were first given by David Cory et al. (1997) and Neil Gershenfeld and Isaac Chuang (1997). Three difficulties had to be overcome for NMR QIP to become possible. The first was that the standard definitions of quantum information and computation require that quantum information be stored in a single physical system. In NMR, an obvious such system consists of some of the nuclear spins

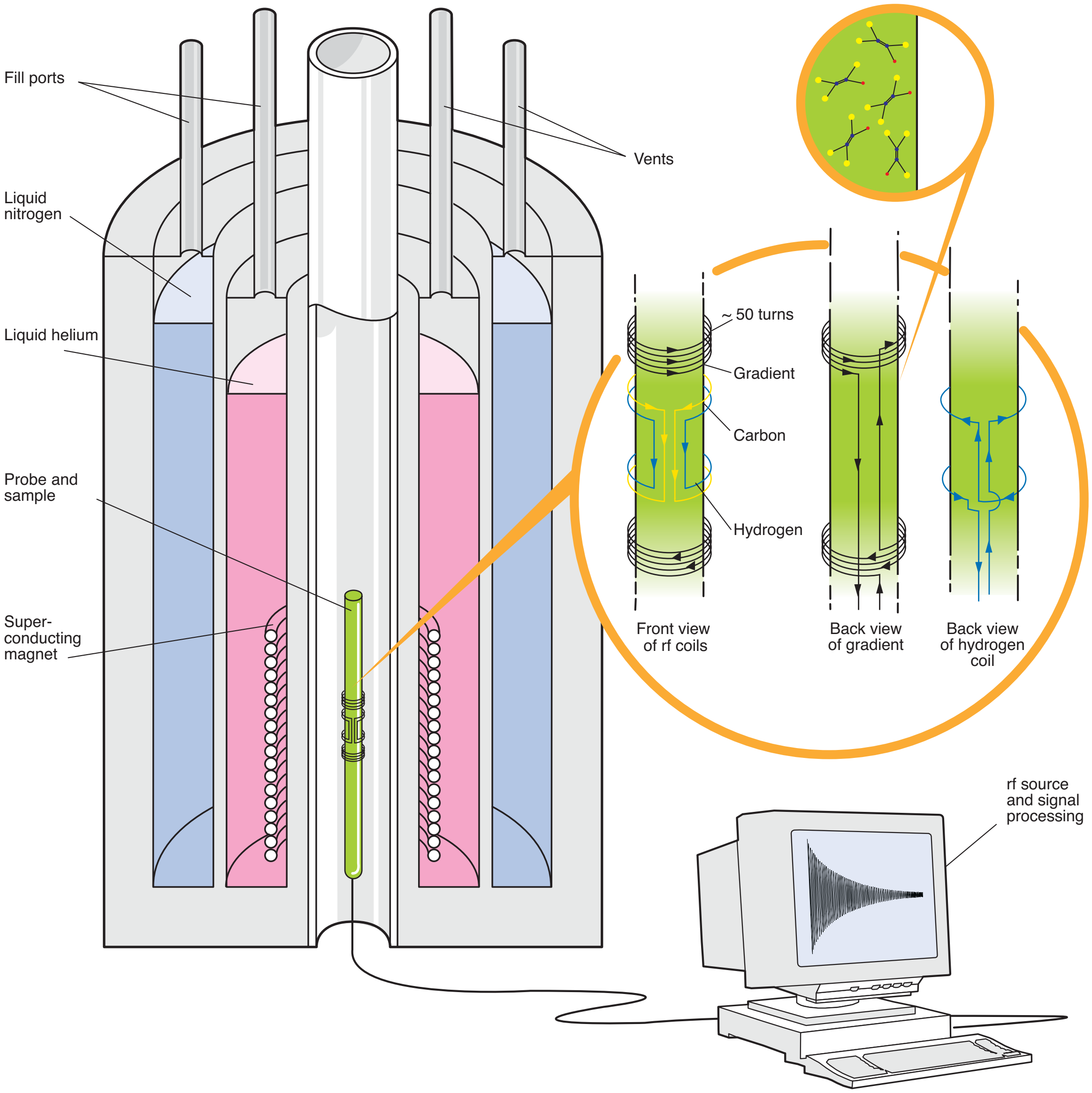

Fill ports
Vents
Liquid nitrogen
Liquid helium
Probe and sample
Super-conducting magnet
~ 50 turns
Gradient
Carbon
Hydrogen
Front view of rf coils
Back view of gradient
Back view of hydrogen coil
rf source and signal processing

Figure 2. Schematic of a Typical NMR Spectrometer (not to scale)
The main components of a spectrometer are the magnet, which is superconducting, and the console , which has the electronics needed to control the spectrometer. The sample containing a liquid solution of the molecule used for QIP is inserted into the central core of the magnet, where it is surrounded by the probe. The probe (shown enlarged to the right) contains coils for applying the rf pulses and magnetic field gradients.

in a single molecule. But it is not possible to detect single molecules with available NMR technology. The solution that makes NMR QIP possible can be applied to other QIP technologies: Consider the large collection of available molecules as an ensemble of identical systems. As long as they all perform the same task, the desired answers can be read out collectively. The second difficulty was that the standard definitions require that readout take place by a projective quantum measurement of the qubits. From such a measurement, one learns whether a qubit is in the state $|0\rangle$ or $|1\rangle$. The two measurement outcomes have probabilities determined by the initial state of the qubits being used, and after the measurement, the state collapses to a state consistent with the outcome.

The measurement in NMR is much too weak to determine the outcome and cause the state's collapse for each molecule. But because of the additive effects of the ensemble, one can observe a (noisy) signal that represents the average, over all the molecules, of the probability that $|1\rangle$ would be the outcome of a projective measurement. It turns out that this so-called weak measurement suffices for realizing most quantum algorithms, in particular those whose ultimate answer is deterministic. Shor's factoring and Lov Grover's search algorithms can be modified to satisfy this property. The final and most severe difficulty was that, even though in equilibrium there is a tendency for the spins to align with the magnetic field, the energy associated with this tendency is very small compared with room temperature. Therefore, the equilibrium states of the molecules' nuclear spins are nearly random, with only a small fraction pointing in the right direction. This difficulty was overcome by methods for singling out the small fraction of the observable signal that represents the desired initial state. These methods were anticipated in 1977 (Stall et al.)

Soon after these difficulties were shown to be overcome or circumventable, two groups were able to experimentally implement short quantum algorithms using NMR with small molecules (Chuang et al. 1998, Jones et al. 1998). At present, it is considered unlikely that liquid-state NMR algorithms will solve problems not easily solvable with available classical computing resources. Nevertheless, experiments in liquid-state NMR QIP are remarkable for demonstrating that one can control the unitary evolution of physical qubits sufficiently well to implement simple QIP tasks. The control methods borrowed from NMR and developed for the more complex experiments in NMR QIP are applicable to other device technologies, enabling better control in general.

Principles of Liquid-State NMR QIP

In order to physically realize quantum information, it is necessary to find ways of representing, manipulating, and coupling qubits so as to implement nontrivial quantum gates, prepare a useful initial state, and read out the answer. The next sections show how to accomplish these tasks in liquid-state NMR.

Realizing Qubits. The first step for implementing QIP is to have a physical system that can carry quantum information. The preferred system for realizing qubits in liquid-state NMR consists of spin-1/2 nuclei, which are naturally equivalent to qubits. The nuclear-spin degree of freedom of a spin-1/2 nucleus defines a quantum mechanical two-state system. Once the direction along the strong external magnetic field is fixed, its state space consists of the superpositions of "up" and "down" states. That is, we can imagine that the nucleus behaves somewhat like a small magnet, with a definite axis, which can point either up (logical state $|0\rangle$) or down (logical state $|1\rangle$). By the superposition principle, every quantum state of the form $|\psi_0\rangle = \alpha|0\rangle + \beta|1\rangle$ with $|\alpha|^2 + |\beta|^2 = 1$ is a possible (pure) state for the nuclear spin. In the external magnetic field, the two logical states have different energies. In quantum mechanics, this observation means that the time evolution of $|\psi_0\rangle$ is given by

$$\left|\psi_t\right\rangle = e^{-i\omega t/2}\alpha\left|0\right\rangle + e^{i\omega t/2}\beta\left|1\right\rangle . \qquad (1)$$

The constant ω is the precession frequency of the nuclear spin in the external magnetic field in units of radians per second if t is in seconds. The frequency is proportional to the energy difference ε between the up and down states: $\omega = 2\pi\varepsilon/h$, where h is Planck's constant.

Although a spin-1/2 nucleus' state space is the same as that of a qubit, the precession implies that the state is not constant. We would like the realization of a qubit to retain its state over time when we are not intentionally modifying it. For this reason, in the next section, the qubit state realized by the nuclear spin will be defined so as to compensate for the precession.

Precession frequencies for nuclear spins can vary substantially depending on the nuclei's magnetic moments. For example, at 11.7 tesla, the precession frequency for protons is 500 megahertz, and for carbon-13, it is 125 megahertz. These frequency differences are exploited in measurement and control to distinguish between the types of nuclei. The effective magnetic field seen by nuclear spins also depends on their chemical environment. This dependence causes small variations in the spins' precession frequencies that can be used to distinguish, for example, the two carbon-13 nuclei in TCE: The frequency difference (called the "chemical shift") is 600 to 900 hertz at 11.7 tesla, depending on the solvent, the temperature, and the TCE concentration.

Pauli Matrices

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

If we use the Pauli matrix σ_z, the time evolution can be expressed as $|\psi_t\rangle = e^{iw\sigma_z t/2}|\psi_0\rangle$. The operator $\omega\sigma_z/2$ is the internal Hamiltonian (that is, the energy observable, in units for which $h/(2\pi) = 1$) of the nuclear spin. The direction of the external magnetic field determines the z-axis. Given a choice of axes, the idea that a single nuclear spin 1/2 has a direction (as would be expected for a tiny magnet) can be made explicit by means of the Bloch sphere representation of a nuclear spin's state (refer to Figure 3). The Pauli matrix σ_z can be thought of as the observable that measures the nuclear spin along the z-axis. Observables for spin along the x- and y-axis are given by the other two Pauli matrices, σ_x and σ_y. Given a state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ of the nuclear spin, one can form the density matrix $|\psi\rangle\langle\psi|$ and express it in the form

$$|\psi\rangle\langle\psi| = \frac{1}{2}\left(\mathbb{1} + \alpha_x\sigma_x + \alpha_y\sigma_y + \alpha_z\sigma_z\right) . \tag{2}$$

The vector $\boldsymbol{v} = (\alpha_x, \alpha_y, \alpha_z)$ then is a point on the unit sphere in three-dimensional space. Conversely, every point on the unit sphere corresponds to a pure state of the nuclear spin. The representation also works for mixed states, which correspond to points in the interior of the sphere. As a representation of spin states, the unit sphere is called the Bloch sphere. Because quantum evolutions of a spin correspond to Bloch sphere rotations, the Bloch sphere is a useful tool for thinking about one- and sometimes about two-qubit processes.

If we write the state as a density matrix ρ and expand it in terms of Pauli matrices,

$$\begin{aligned} \rho &= |\psi\rangle\langle\psi| = \left(\mathbb{1} + x\sigma_x + y\sigma_y + z\sigma_z\right)\big/2 \\ &= \frac{1}{2}\left(\mathbb{1} + \sin(\theta)\cos(\phi)\sigma_x + \sin(\theta)\sin(\phi)\sigma_y + \cos(\theta)\sigma_z\right) , \end{aligned} \tag{3}$$

the coefficients $(x, y, z) = (\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta))$ of the Pauli matrices form the vector for the state. The angles θ and ϕ are the Euler angles, as shown in Figure 3. For a pure state, this vector is on the surface of the unit sphere, and for a mixed state, it is inside the unit sphere. The Pauli matrices are associated with spin observables in the laboratory frame, so that all axes of the representation are meaningful with respect to real space.

One-Qubit Gates. The second step for realizing QIP is to give a means for controlling the qubits so that quantum algorithms can be implemented. The qubits are controlled with carefully modulated external fields to realize specific unitary evolutions called gates.

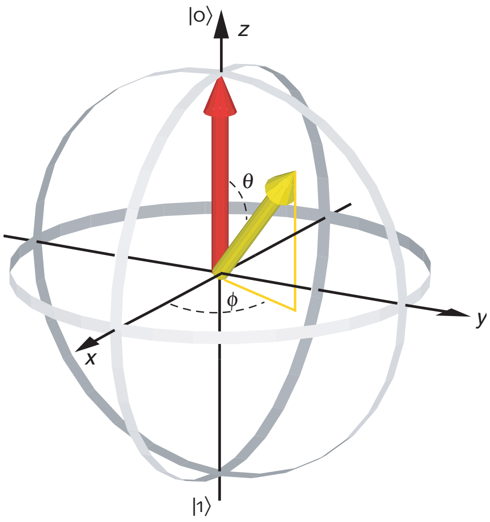


Figure 3. Bloch Sphere Representation of a Qubit State

The yellow arrow represents a pure state $|\psi\rangle$ for the qubit or the nuclear spin 1/2. The Euler angles are indicated and determine the state according to the formula $|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\phi/2)|1\rangle$. The red arrow along the *z*-axis indicates the orientation of the magnetic field and the vector for $|0\rangle$.

Each such evolution can be described by a unitary operator applied to one or more qubits. The simplest method for demonstrating that sufficient control is available is to show how to realize a set of one- and two-qubit gates that is universal in the sense that, in principle, every unitary operator can be implemented as a composition of gates (Barenco et al. 1995, DiVincenzo 1995, Lloyd 1995).

One-qubit gates can be thought of as rotations of the Bloch sphere and can be implemented in NMR with electromagnetic pulses. In general, the effect of a magnetic field on a nuclear spin is to cause a rotation around the direction of the field. In terms of the quantum state of the spin, the effect is described by an internal Hamiltonian of the form $H = (\omega_x\sigma_x + \omega_y\sigma_y + \omega_z\sigma_z)/2$. The coefficients of the Pauli matrices depend on the magnetic field according to $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z) = -\mu\mathbf{B}$, where μ is the nuclear magnetic moment and $\mathbf{B}$ is the magnetic field vector. In terms of the Hamiltonian, the evolution of the spin's quantum state in the presence of the magnetic field $\mathbf{B}$ is therefore given by $|\psi_t\rangle = e^{-iHt}|\psi_0\rangle$ so that the spin direction in the Bloch sphere rotates around $\boldsymbol{\omega}$ with angular frequency $\omega = |\boldsymbol{\omega}|$.

In the case of liquid-state NMR, there is an external, strong magnetic field along the *z*-axis, and the applied electromagnetic pulses add to this field. One can think of these pulses as contributing a relatively weak magnetic field (typically less than .001 of the external field), whose orientation is in the *xy*-plane. One use of such a pulse is to tip the nuclear spin from the *z*-axis to the *xy*-plane. To see how that can be done, assume that the spin starts in the state $|0\rangle$, which points up along the *z*-axis in the Bloch sphere representation. Because this state is aligned with the external field, it does not precess. To tip the spin, one can start by applying a pulse field along the *x*-axis. Because the pulse field is weak compared with the external field, the net field is still almost along the *z*-axis. The spin now rotates around the net field. Since it started along *z*, it moves only in a small circle near the *z*-axis. To force the spin to tip further, one changes the orientation of the pulse field at the same frequency as the precession frequency. This is called a resonant pulse. Because typical precession frequencies are hundreds of megahertz, such a pulse consists of rf electromagnetic fields.

To better understand how resonant pulses work, it is convenient to use the "rotating frame." In this frame, we imagine that our apparatus rotates at the precession frequency of the nuclear spin. In this way, the effect of the external field is removed. In particular, in the rotating frame, the nuclear spin does not precess, and a resonant pulse's magnetic field looks like a constant magnetic field applied, for example, along the $(-x)$-axis of the rotating frame. The nuclear spin responds to the pulse by rotating around the *x*-axis, as expected: If the spin starts along the *z*-axis, it tips toward the $(-y)$-axis, then goes to

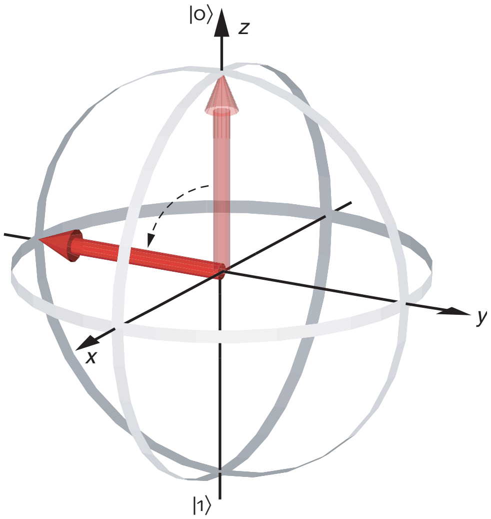


Figure 4. Single-Bit Rotation around the *x*-Axis in the Rotating Frame
An applied magnetic field along the rotating frame's (–*x*)-axis due to a resonant rf pulse moves the nuclear spin direction from the *z*- toward the (–*y*)-axis. The initial and final states for the nuclear spin are shown for a 90° rotation. If the strength of the applied magnetic field is such that the spin evolves according to the Hamiltonian $\omega_x\sigma_x/2$, then it has to be turned on for a time $t = \pi/(2\omega_x)$ to cause the rotation shown.

the (–z)-, to the y-, and finally back to the z-axis, all in the rotating frame (see Figure 4).

The rotating frame makes it possible to define the state of the qubit realized by a nuclear spin as the state with respect to this frame. As a result, the qubit's state does not change unless rf pulses are applied. In the context of the qubit realized by a nuclear spin, the rotating frame is called the logical frame. In the following, references to the Bloch sphere axes and associated observables are understood to be with respect to an appropriate, usually rotating, frame. Different frames can be chosen for each nuclear spin of interest, so we often use multiple independently rotating frames and refer each spin's state to the appropriate frame.

Use of the rotating frame together with rf pulses makes it possible to implement all one-qubit gates on a qubit realized by a spin-1/2 nucleus. To apply a rotation around the x-axis, a resonant rf pulse with effective field along the rotating frame's (–x)-axis is applied. This is called an x-pulse, and x is the "axis" of the pulse. While the rf pulse is on, the qubit's state evolves as $e^{-i\omega_x\sigma_x t/2}$. The strength (or power) of the pulse is characterized by ω_x, the nutation frequency. To implement a rotation by an angle of ϕ, the pulse is turned on for a period $t = \phi/\omega_x$. Rotations around any axis in the plane can be implemented similarly. The angle of the pulse field with respect to the (–x)-axis is called the phase of the pulse. It is a fact that all rotations of the Bloch sphere can be decomposed into rotations around axes in the plane. For rotations around the z-axis, an easier technique is possible. The current absolute phase θ of the rotating frame's x-axis is given by $\theta_0 + \omega t$, where ω is the precession frequency of the nuclear spin. Changing the angle θ_0 by $-\phi$ is equivalent to rotating the qubit's state by ϕ around the z-axis. In this sense, z-pulses can be implemented exactly. In practice, this change of the rotating frame's phase means that the absolute phases of future pulses must be shifted accordingly. This implementation of rotations around the z-axis is possible because phase control in modern equipment is extremely reliable so that errors in the phase of applied pulses are negligible compared with other sources of errors.

So far, we have considered just one nuclear spin in a molecule. But the rf fields are experienced by the other nuclear spins as well. This side effect is a problem if only one target nuclear spin's state is to be rotated. There are two cases to consider depending on the precession frequencies of the other, nontarget spins. Spins of nuclei of different isotopes, such as those of other species of atoms, usually have precession frequencies that differ from the target's by many megahertz at 11.7 tesla. A pulse resonant for the target has little effect on such spins because, in the rotating frames of the nontarget spins, the pulse's magnetic field is not constant but rotates rapidly. The power of a typical pulse is such that the effect during one rotation of the pulse's field direction is insignificant and averages to zero over many rotations. This is not the case for nontarget spins of the same isotope. Although the variations in their chemical environments result in frequency differences, these differences are much smaller, often only a few kilohertz. The period of a 1-kilohertz rotation is 1 millisecond, whereas so-called hard rf pulses require only tens of microseconds (.00l millisecond) to complete the typical 90° or 180° rotations. Consequently, in the rotating frame of a nontarget spin with a small frequency difference, a hard rf pulse's magnetic field is nearly constant for the duration of the pulse. As a result, such a spin experiences a rotation similar to the one intended for the target. To rotate a specific nuclear spin or spins within a narrow range of precession frequencies, one can use weaker, longer-lasting "soft" pulses instead. This approach leads to the following strategies for applying pulses: To rotate all the nuclear spins of a given species (such as the two carbon-13 nuclei of TCE) by a desired angle, apply a hard rf pulse for as short a time as possible. To rotate just one spin having a distinct precession frequency, apply a soft rf pulse of sufficient duration to have little effect on other spins. The power of soft pulses is usually modulated in time ("shaped") to reduce the time needed for a rotation while minimizing crosstalk, a term that describes unintended effects on other nuclear spins.

Two-Qubit Gates. Two nuclear spins in a molecule interact with each other, as one would expect of two magnets. But the details of the spins' interaction are more complicated because they are mediated by the electrons. In liquid state, the interaction is also modulated by the rapid motions of the molecule. The resulting effective interaction is called the J-coupling. When the difference of the precession frequencies between the coupled nuclear spins is large compared with the strength of the coupling, it is a good approximation to write the coupling Hamiltonian as a product of the z-Pauli operators for each spin: $H_J = C\sigma_z^{(1)}\sigma_z^{(2)}$. This is the weak-coupling regime. With this Hamiltonian, an initial state $|\psi_0\rangle$ of two nuclear-spin qubits evolves as $|\psi_t\rangle = e^{-iC\sigma_z^{(1)}\sigma_z^{(2)}t}|\psi_0\rangle$, where a different rotating frame is used for each nuclear spin to eliminate the spin's internal evolution. (The use of rotating frames is compatible with the coupling Hamiltonian because the Hamiltonian is invariant under frame rotations.) Because the Hamiltonian is diagonal in the logical basis, the effect of the coupling can be understood as an increase of the (signed) precession frequency of the second spin if the first one is up and a decrease if the first one is down (see Figure 5). The changes in precession frequency for adjacent nuclear spins in organic molecules are typically in the range of 20 to 200 hertz. They are normally much smaller for nonadjacent nuclear spins. The strength of the coupling is called the coupling constant and is given as the change in the precession frequency. In terms of the constant C used above, the coupling constant is given by $J = 2C/\pi$ in hertz. For example, the coupling constants in TCE are close to 100 hertz between the two carbons, 200 hertz between the proton and the adjacent carbon, and 9 hertz between the proton and the far carbon.

The J-coupling and the one-qubit pulses suffice for realizing the controlled-not operation usually taken as one of the fundamental gates of QIP. A pulse sequence for implementing the controlled-not in terms of the J-coupling constitutes the first quantum algorithm discussed under "Examples of Quantum Algorithms for NMR." A problem with the J-coupling in liquid-state NMR is that it cannot be turned off when it is not needed for implementing a gate.

Turning off the *J*-Coupling. The coupling between the nuclear spins in a molecule cannot be physically turned off. But for QIP, we need to be able to maintain a state in memory and to couple qubits selectively. Fortunately, NMR spectroscopists solved this problem well before the development of modern quantum-information concepts. The idea is to use the control of single spins to cancel the interaction's effect over a given

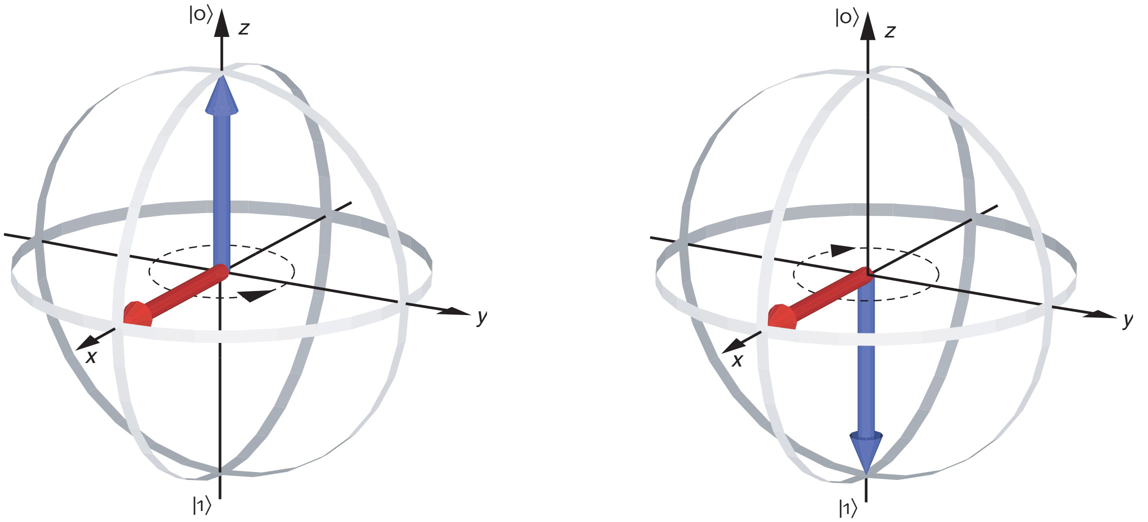


Figure 5. *J*-Coupling Effect
In the weak-coupling regime with a positive coupling constant, the coupling between two spins can be interpreted as an increase in precession frequency of spin 2 when spin 1 is up and a decrease when spin 1 is down. The two diagrams depict the situation in which spin 2 is in the plane. The diagram on the left has spin 1 pointing up along the *z*-axis. In the rotating frame of spin 2, it precesses from the *x*-axis to the *y*-axis. The diagram on the right has spin 1 pointing down, causing a precession in the opposite direction of spin 2. Note that neither the coupling nor the external field changes the orientation of a spin pointing up or down along the *z*-axis.

period. This technique is called refocusing and requires applying a 180° pulse to one of two coupled spins at the midpoint of the desired period. To understand how refocusing works, consider again the visualization of Figure 5. A general state is in a superposition of the four logical states of the two spins. By linearity, it suffices to consider the evolution with spin 1 being in one of its two logical states, up or down, along the z-axis. Suppose we wish to remove the effects of the coupling over a period of 2 milliseconds. To do so, wait 1 millisecond. In a sequence of pulses, this waiting period is called a 1-millisecond delay. The effect on spin 2 in its rotating frame is to precess counterclockwise if spin 1 is up and clockwise for the same angle if spin 1 is down. Now, apply a pulse that rotates spin 1 by 180° around the x-axis. This is called an inversion, or in the current context, a refocusing pulse. It exchanges the up and down states. For the next 1 millisecond, the effect of the coupling on spin 2 is to undo the earlier rotation. At the end of the second 1-millisecond delay, one can apply another 180° pulse to reverse the inversion and recover the initial state. The pulse sequence is depicted in Figure 6.

Turning off couplings between more than two nuclear spins can be quite complicated unless one takes advantage of the fact that nonadjacent nuclear spins tend to be relatively weakly coupled. Methods that scale polynomially with the number of nuclear spins and that can be used to selectively couple pairs of nuclear spins can be found in Debbie Leung et al. (1999) and Jonathan Jones and Knill (1999). These techniques can be used in other physical systems, where couplings exist that are difficult to turn off directly. An example is qubits represented by the state of one or more electrons in tightly packed quantum dots.

Measurement. To determine the answer of a quantum computation, it is necessary to make a measurement. As noted earlier, the technology for making a projective measurement of individual nuclear spins does not yet exist. In liquid-state NMR, instead of using just one molecule to define a single quantum register, we use a large ensemble of molecules in a test tube. Ideally, their nuclear spins are all placed in the same initial state, and the subsequent rf pulses affect each molecule in the same way. As a result, weak magnetic signals from, say, the proton spins in TCE add to form a detectable magnetic field called the bulk magnetization. The signal that is measured in high-field NMR is the magnetization in the xy-plane, which can be picked up by coils whose axes are placed transversely to the external field. Because the interaction of any given nuclear spin with the coil is very weak, the effect of the coil on the quantum state of the spins is negligible in most NMR experiments. As a result, it is a good approximation to think of the generated magnetic fields and their detection classically. In this approximation, each nuclear spin behaves like a tiny bar magnet and contributes to the bulk magnetization.

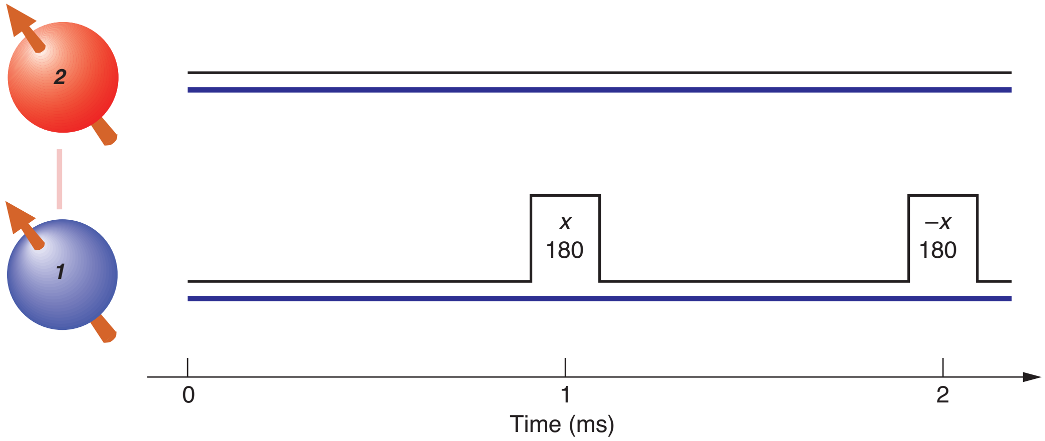


Figure 6. Pulse Sequence for Refocusing the Coupling

The sequence of events is shown with time running from left to right. The two spins' lifelines are shown in blue, and the rf power targeted at each spin is indicated by the black line above. Pulses are applied to spin 1 only, as indicated by the rectangular rises in rf power at 1 ms and 2 ms. The axis for each pulse is given with the pulse. The angle is determined by the area under the pulse and is also given explicitly. Ideally, for pulses of this type, the pulse times (the widths of the rectangles) should be zero. In practice, for hard pulses, they can be as small as ≈ .01 ms. Any $\sigma_z^{(1)}\sigma_z^{(2)}$ coupling's effect is refocused by the sequence shown so that the final state of the two spins is the same as the initial state. The axis for the pair of refocusing pulses can be changed to any other axis in the plane.

As the nuclear spins precess, so does the magnetization. As a result, an oscillating current is induced in the coil, provided it is electronically configured to be tuned to the precession frequency. By observing the amplitude and phase of this current over time, we can keep track of the absolute magnetization in the plane and its phase with respect to the rotating frame. This process yields information about the qubit states represented by the state of the nuclear spins.

To see how one can use bulk magnetization to learn about the qubit states, consider the TCE molecule with three spin-1/2 nuclei used for information processing. The bulk magnetizations generated by the protons and the carbons precess at 500 megahertz and 125 megahertz, respectively. The proton and carbon contributions to the magnetization are detected separately with two coils tuned to 500 megahertz (proton magnetization) and 125 megahertz (carbon magnetization). For simplicity, we restrict our attention to the two carbons and assume that the protons are not interacting with the carbons. (It is possible to actively remove such interactions by using a technique called decoupling.)

At the end of a computation, the qubit state of the two nuclear spins is given by a density matrix ρ_q. We can assume that this state is the same for each TCE molecule in the sample. As we mentioned earlier, the density matrix is relative to logical frames for each nuclear spin. The current phases for the two logical frames with respect to a rotating reference frame at the precession frequency of the first carbon are known. If we learn something about the state in the reference frame, that information can be converted to the desired logical frame by a rotation around the z-axis. Let $\rho(0)$ be the state of the two nuclear spins in the reference frame. In this frame, the state evolves in time as $\rho(t)$ according to a Hamiltonian H that consists of a chemical shift term for the difference in the precession frequency of the second carbon and of a coupling term. To a good approximation,

$$H = \pi 900\mathrm{Hz}\,\sigma_z^{(2)} + \pi 50\mathrm{Hz}\,\sigma_z^{(1)}\sigma_z^{(2)} \quad . \tag{4}$$

The magnetization detected in the reference x-direction at time t is given by

$$M_x(t) = m\,\mathrm{tr}\Big(\rho(t)\Big(\sigma_x^{(1)} + \sigma_x^{(2)}\Big)\Big) \quad , \tag{5}$$

where tr(σ) denotes the trace, that is, the sum of the diagonal elements of the matrix σ. Equation 5 links the magnetization to the Bloch sphere representation. The constant of proportionality m depends on the size of the ensemble and the magnetic moments of the nuclei. From the point of view of NMR, m determines a scale whose absolute size is not relevant. What matters is how strong this signal is compared with the noise in the system. For the purpose of the following discussion, we set $m = 1$.

We can also detect the magnetization $M_y(t)$ in the y-direction and use this result together with $M_x(t)$ to form a complex number representing the planar magnetization.

$$\begin{aligned} M(t) &= M_x(t) + iM_y(t) && (6)\\ &= \operatorname{tr}\Big(\rho(t)\big(\sigma_+^{(1)} + \sigma_+^{(2)}\big)\Big)\ , && (7) \end{aligned}$$

where we defined $\sigma_+ = \sigma_x + i\sigma_y = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$.

What can we infer about $\rho(0)$ from observing $M(t)$ over time? For the moment, we neglect the coupling Hamiltonian. Under the chemical shift Hamiltonian $H_{\mathrm{CS}} = \pi 900\mathrm{Hz}\sigma_z^{(2)}$, $M(t)$ evolves as

$$\begin{aligned} M(t) &= \operatorname{tr}\Big(e^{-iH_{CS}t}\rho(0)e^{iH_{CS}t}\big(\sigma_+^{(1)} + \sigma_+^{(2)}\big)\Big) \\ &= \operatorname{tr}\Big(\rho(0)e^{iH_{CS}t}\big(\sigma_+^{(1)} + \sigma_+^{(2)}\big)e^{-iH_{CS}t}\Big) && \text{Use } \operatorname{tr}(AB) = \operatorname{tr}(BA). \\ &= \operatorname{tr}\Big(\rho(0)\big(\sigma_+^{(1)} + e^{iH_{CS}t}\sigma_+^{(2)}e^{-iH_{CS}t}\big)\Big) && H_{CS} \text{ acts only on spin 2.} \\ &= \operatorname{tr}\Big(\rho(0)\big(\sigma_+^{(1)} + e^{i2\pi 900\mathrm{Hz}\,t}\sigma_+^{(2)}\big)\Big) && \text{Multiply the matrices.} \\ &= \operatorname{tr}\Big(\rho(0)\sigma_+^{(1)}\Big) + \operatorname{tr}\Big(\rho(0)e^{i2\pi 900\mathrm{Hz}t}\sigma_+^{(2)}\Big)\ . && \text{The trace is linear.} \quad (8) \end{aligned}$$

Thus, the signal is a combination of a constant signal given by the first spin's contribution to the magnetization in the plane and a signal oscillating with a frequency of 900 hertz with amplitude given by the second spin's contribution to the planar magnetization. The two contributions can be separated by Fourier-transforming $M(t)$, which results in two distinct peaks, one at 0 hertz and a second at 900 hertz (refer to Figure 7).

To see how the coupling affects the observed magnetization, we rewrite the expression for $M(t)$ to take advantage of the fact that the up-down states are invariant under the full Hamiltonian.

$$\begin{aligned} M(t) &= \operatorname{tr}\Big(\rho(t)\sigma_+^{(1)}\Big) + \operatorname{tr}\Big(\rho(t)\sigma_+^{(2)}\Big) \\ &= \operatorname{tr}\Big(\rho(t)\sigma_+^{(1)}\mathbb{1}^{(2)}\Big) + \operatorname{tr}\Big(\rho(t)\,\mathbb{1}^{(1)}\sigma_+^{(2)}\Big) \\ &= \operatorname{tr}\Big(\rho(t)\sigma_+^{(1)}\big(e_\uparrow^{(2)} + e_\downarrow^{(2)}\big)\Big) + \operatorname{tr}\Big(\rho(t)\big(e_\uparrow^{(1)} + e_\downarrow^{(1)}\big)\sigma_+^{(2)}\Big)\ , && (9) \end{aligned}$$

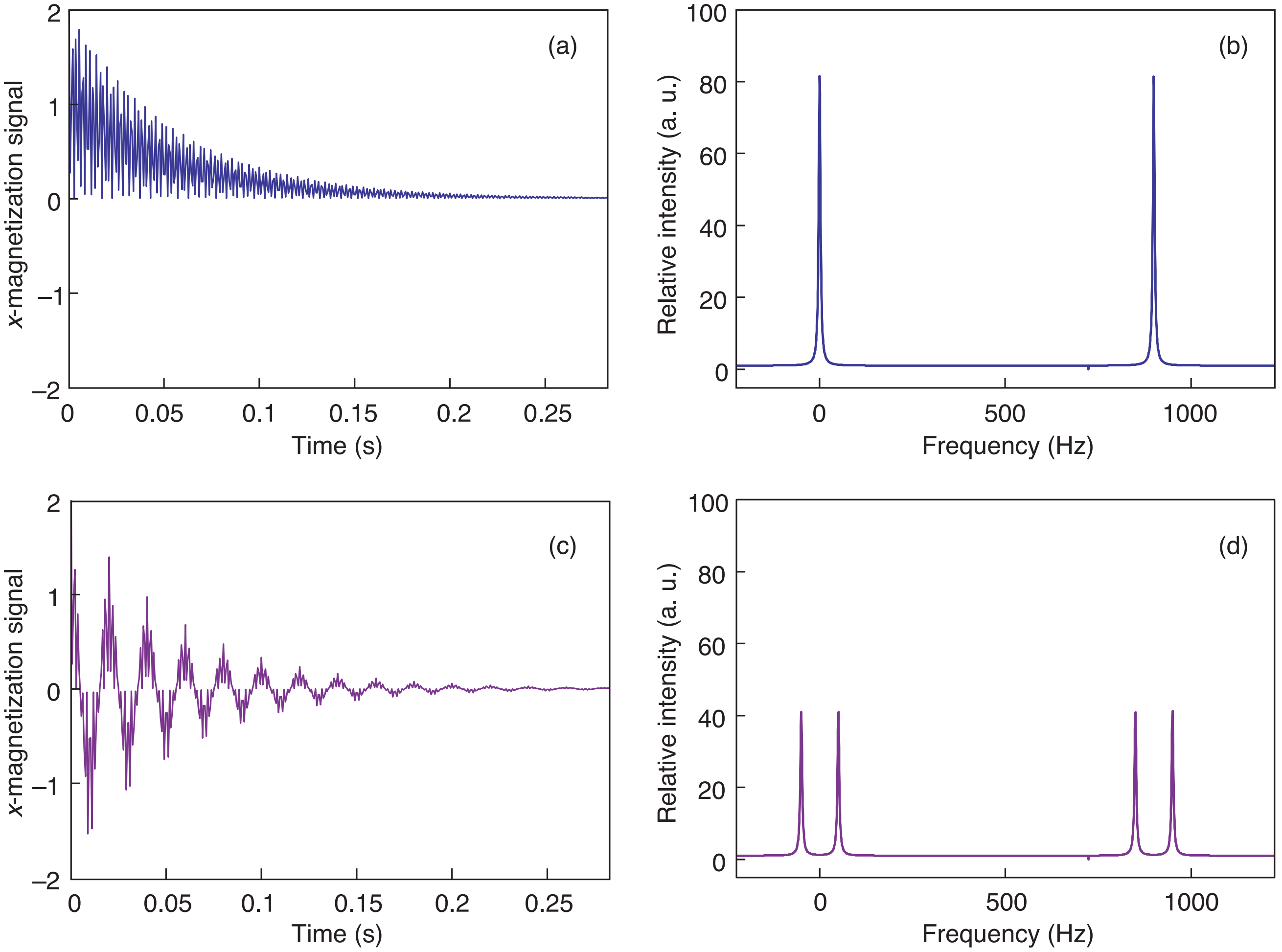


Figure 7. Simulated Magnetization Signals and Spectra

(a) The *x*-magnetization signal is shown as a function of time for a pair of uncoupled spins with a relative chemical shift of 900 Hz. The initial spin directions are along the *x*-axis. The signal (called the "free-induction decay") decays with a halftime of 0.0385 *s* because of simulated relaxation processes. Typically, the halftimes are much longer. A short one was chosen in order to broaden the peaks for visual effect. (b) The spectrum, that is, the Fourier transform of the combined *x*- and *y*-magnetization has peaks at frequencies of 0 Hz (spin 1's peak) and 900 Hz (spin 2's peak) because of the independently precessing pair of spins. (c) This plot shows the *x*-magnetization signal when the two spins coupled as described in the text. (d) Shown here is the spectrum for the signal in (c) obtained from combined *x*- and *y*-magnetization. Each spin's peak from the previous spectrum "splits" into two. The left and right peaks of each pair are associated with the other spin being in the state |1⟩ and |0⟩, respectively. The vertical axis units are relative intensity with the same constant of proportionality for the two spectra.

where $e_{\uparrow} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $e_{\downarrow} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

Using a calculation similar to the one leading to Equation (8), the first term can be written as

$$M_1(t) = \mathrm{tr}\left(e^{-iHt}\rho(0)e^{iHt}\sigma_+^{(1)}\left(e_{\uparrow}^{(2)}+e_{\downarrow}^{(2)}\right)\right) \qquad (10)$$

$$= e^{i2\pi 50\mathrm{Hz}t}\,\mathrm{tr}\left(\rho(0)\,\sigma_+^{(1)}e_{\uparrow}^{(2)}\right) + e^{-i2\pi 50\mathrm{Hz}\,t}\,\mathrm{tr}\left(\rho(0)\sigma_+^{(1)}e_{\downarrow}^{(2)}\right) , \qquad (11)$$

and similarly for the second term, but with an offset frequency of 900 hertz because of the chemical shift. It can be seen that the zero-frequency signal splits into two signals with frequencies of –50 hertz and 50 hertz, respectively. The difference between the two frequencies is the coupling constant. The amplitudes of the different frequency signals can be used to infer the expectations of operators such as $\sigma_+{}^{(1)}e_\uparrow{}^{(2)}$, given by tr $(r(0)\sigma_+{}^{(1)}e_\uparrow{}^{(2)})$. For n spin-1/2 nuclei, the spectral peak of a nucleus splits into a group of 2^{n-1} peaks, each associated with operators such as $\sigma_+{}^{(a)}e_\uparrow{}^{(b)}e_\downarrow{}^{(c)}e_\downarrow{}^{(d)}\ldots$ Later in the article (see figure on page 249), we show a simulated peak group for a nuclear spin coupled to three other spins. Expectations of the single-spin operators $\sigma_x{}^{(a)}$ and $\sigma_y{}^{(a)}$ can be obtained from the real and imaginary parts of the total signal in a peak group for a nucleus. The positions of the 2^{n-1} peaks depend on the couplings. If the peaks are well separated, we can infer expectations of product operators with only one σ_x or σ_y, such as $\sigma_x{}^{(a)}\sigma_z{}^{(b)}\mathbb{1}^{(c)}\sigma_z{}^{(d)}$, by taking linear combinations with appropriate coefficients of the peak amplitudes in a peak group.

In addition to the unitary evolution due to the internal Hamiltonian, relaxation processes tend to decay $\rho(t)$ toward the equilibrium state. In liquid state, the equilibrium state ρ_{thermal} is close to $\mathbb{1}/N$, where N is the total dimension of the state space. The difference between ρ_{thermal} and $\mathbb{1}/N$ is the equilibrium "deviation" density matrix and has magnetization only along the z-axis (see the section "The Initial State"). Because the only observed magnetization is planar, the observed signal decays to zero as the state relaxes to equilibrium. To a good approximation, we can write

$$\rho(t) = \frac{1}{N}\mathbb{1} + e^{-\lambda t}\rho'(t) + (\text{not observed}) \ , \tag{12}$$

where $\rho'(t)$ has trace zero and evolves unitarily under the Hamiltonian. The effect of the relaxation process is that $M(t)$ has an exponentially decaying envelope, explaining the conventional name for $M(t)$, namely, the free induction decay (FID). Typical halftimes for the decay are .1 to 2 seconds for nuclear spins used for QIP. A normal NMR observation consists of measuring $M(t)$ at discrete time intervals until the signal is too small. The acquired FID is then Fourier-transformed to visualize the amplitudes of the different frequency contributions. The shape of the peaks in Figure 7 reflects the decay envelope. The width of the peaks is proportional to the decay rate λ.

For QIP, we wish to measure the probability p that a given qubit (say, qubit 1) is in the state $|1\rangle_1$. We have $1 - 2p = \mathrm{tr}(\rho\sigma_z{}^{(1)})$, which is the expectation of $\sigma_z{}^{(1)}$. We can measure this expectation by first applying a 90° y-pulse to qubit 1 and thus changing the state to ρ'. This pulse has the effect of rotating initial, unobservable z-magnetization to observable x-magnetization. From $M(t)$ one can then infer $\mathrm{tr}(\rho'\sigma_x{}^{(1)})$, which is the desired number. For the coupled pair of carbons, $\mathrm{tr}(\rho'\sigma_x{}^{(1)})$ is given by the sum of the real components of the amplitudes of the 50 hertz and the –50 hertz contributions to $M(t)$. However, the problem is that these amplitudes are determined only up to a scale. A second problem is that the available states ρ are highly mixed (close to $\mathbb{1}/N$). The next section discusses how to compensate for both problems.

As a final comment on NMR measurement, note that the back reaction on the nuclear spins due to the emission of electromagnetic energy is weak. This is what enables us to measure the bulk magnetization over some time. The ensemble nature of the system gives direct, if noisy, access to expectations of observables such as σ_z rather than a single answer—0 or 1. For algorithms that provide a definite answer, having access only to expectations is not a problem because it is easy to distinguish the answer from the noise. However, using expectations can increase the need for quantum resources. For example, Shor's factoring algorithm includes a significant amount of classical postprocessing based on highly random answers from projective measurements. In order to implement the algorithm in an ensemble setting, the postprocessing must be performed reversibly and integrated into the quantum computation to guarantee a definite answer. Postprocessing can be done with polynomial additional quantum resources.

The Initial State. Because the energy difference between the nuclear spins' up and down states is so small compared with room temperature, the equilibrium distribution of states is nearly random. In the liquid samples used, equilibrium is established after l0 to 40 seconds if no rf fields are being applied. As a result, all computations start with the sample in equilibrium. One way to think of this initial state is that every nuclear spin in each molecule begins in the highly mixed state $(1 - \varepsilon)\mathbb{1}/2 + \varepsilon|0\rangle\langle 0|$, where ε is a small number (of the order of 10^{-5}). This is a nearly random state with a small excess of the state $|0\rangle$. The expression for the initial state derives from the fact that the equilibrium state ρ_{thermal} is proportional to $e^{-H/kT}$, where H is the internal Hamiltonian of the nuclear spins in a molecule (in energy units), T is the temperature, and k is the Boltzman constant. In our case, H/kT is very small, and the coupling terms are negligible. Therefore,

$$e^{-H/kT} \approx e^{-\varepsilon_1\sigma_z^{(1)}/kT}\, e^{-\varepsilon_2\sigma_z^{(2)}/kT} \ldots, \tag{13}$$

$$e^{-\varepsilon_1\sigma_z^{(1)}/kT} \approx \mathbb{1} - \varepsilon_1\sigma_z^{(1)}/kT, \text{ and} \tag{14}$$

$$e^{-H/kT} \approx \mathbb{1} - \varepsilon_1\sigma_z^{(1)}/kT - \varepsilon_2\sigma_z^{(2)}/kT - \ldots, \tag{15}$$

where ε_l is half of the energy difference between the up and down states of the lth nuclear spin.

Clearly, the available initial state is very far from what is needed for standard QIP. However, it can still be used to perform interesting computations. The main technique is to use available NMR tools to change the initial state to a pseudopure state, which for all practical purposes, behaves like the initial state required by QIP. The technique is based on three key observations. First, only the traceless part of the density matrix contributes to the magnetization. Suppose that we are using n spin-1/2 nuclei in a molecule and the density matrix is ρ. Then, the current magnetization is proportional to $\mathrm{tr}(\rho\,\hat{m})$**,** where $\hat{m}$ is a traceless operator—see Equation (9). Therefore, the magnetization does not depend on the part of ρ proportional to the identity matrix. A deviation density matrix for ρ is any matrix δ such that $\delta - \rho = \lambda\mathbb{1}$ for some λ. For example, $\varepsilon|0\rangle\langle 0|$ is a deviation for the equilibrium state of one nuclear spin. We have

$$\begin{aligned}\mathrm{tr}\left(\delta\hat{m}\right) &= \mathrm{tr}\left(\left(\rho + \lambda\mathbb{1}\right)\hat{m}\right)\\ &= \mathrm{tr}\left(\rho\hat{m}\right) + \mathrm{tr}\left(\hat{m}\right)\\ &= \mathrm{tr}\left(\rho\hat{m}\right)\ .\end{aligned} \tag{16}$$

The second observation is that all the unitary operations used, as well as the nonunitary ones to be discussed below, preserve the completely mixed state $\mathbb{1}/2^n$.[1] Therefore, all future observations of magnetization depend only on the initial deviation.

The third observation is that all the scales are relative. In particular, as will be explained, the probability that the final answer of a quantum computation is 1 can be expressed as the ratio of two magnetizations. It follows that one can arbitrarily rescale a deviation density matrix. For measurement, the absolute size of the magnetizations is not important; the most important issue is that the magnetizations are strong enough to be observable over the noise.

---

[1] The intrinsic relaxation process does not preserve the completely mixed state. But its contribution is either negligible over the time scale of typical experiments or can be removed with the help of subtractive phase cycling.

To explain the relativity of the scales and introduce pseudopure states for QIP, we begin with one spin-1/2 qubit. Its equilibrium state has a deviation $\delta = \varepsilon|0\rangle\langle 0|$. If U is the total unitary operator associated with a computation, then δ is transformed to $\delta = \varepsilon U|0\rangle\langle 0|U^\dagger$. For QIP purposes, the goal is to determine what the final probability ρ_1 of measuring $|1\rangle$ is, given that $|0\rangle$ is the initial state. This probability can be computed as follows:

$$
\begin{aligned}
p_1 &= \langle 1|U|0\rangle\langle 0|U^\dagger|1\rangle \\
&= \mathrm{tr}\left(U|0\rangle\langle 0|U^\dagger|1\rangle\langle 1|\right) && (17) \\
&= \mathrm{tr}\left(U|0\rangle\langle 0|U^\dagger\left(\mathbb{1} - \sigma_z\right)\right)\Big/2 && (18) \\
&= \left(\mathrm{tr}\left(U|0\rangle\langle 0|U^\dagger\right) - \mathrm{tr}\left(U|0\rangle\langle 0|U^\dagger\sigma_z\right)\right)\Big/2 && (19) \\
&= \left(1 - \mathrm{tr}\left(U|0\rangle\langle 0|U^\dagger\sigma_z\right)\right)\Big/2 \ . && (20)
\end{aligned}
$$

Thus, the probability can be determined from the expectations of σ_z being measured for the initial and final states (in different experiments). This measurement yields the quantities $a = \mathrm{tr}(\delta\sigma_z) = \varepsilon$ and $a' = \mathrm{tr}(\delta'\sigma_z)\varepsilon\,\mathrm{tr}(U|0\rangle\langle 0|U^\dagger\sigma_z)$, respectively. The desired answer is $p_1 = (1 - (a/a'))\ /2$ and does not depend on the scale ε.

The method presented in the previous paragraph for determining the probability that the answer of a quantum computation is 1 generalizes to many qubits. The goal is to determine the probability p_1 of measuring $|1\rangle_1$ in a measurement of the first qubit after a computation with initial state $|0\ldots 0\rangle$. Suppose we can prepare the spins in an initial state with a deviation $\delta = \varepsilon|0\ldots 0\rangle\langle 0\ldots 0|$. A measurement of the expectations a and a' of σ_z^1 for the initial and final states then yields p_1, as before, by the formula $p_1 = (1 - (a/a'))/2$.

A state with deviation $\varepsilon|\psi\rangle\langle\psi|$ is called a pseudopure state because that deviation is proportional to the deviation of the pure state $|\psi\rangle\langle\psi|$. With respect to scale-independent NMR observations and unitary evolution, a pseudopure state is equivalent to the corresponding pure state. Because NMR QIP methods are scale independent, we now generalize the definition of deviation density matrix: δ is a deviation of the density matrix ρ if $\varepsilon\delta = \rho + \lambda\mathbb{1}$ for some λ and ε.

Among the most important enabling techniques in NMR QIP are the methods that can be used to transform the initial thermal equilibrium state to a standard pseudopure state with deviation $|0\ldots 0\rangle\langle 0\ldots 0|$. An example of how that can be done will be given as the second algorithm in the section "Examples of Quantum Algorithms for NMR." The basic principle for each method is to create, directly or indirectly by summing over multiple experiments, a new initial state as a sum $\rho_0 = \Sigma_i U_i \rho_{\mathrm{thermal}} U_i^\dagger$, where the U_i are carefully and sometimes randomly chosen (Cory et al. 1997, Gershenfeld and Chuang 1997, Knill et al. 1998, Sharf et al. 2000) to ensure that ρ_0 has a standard pseudopure deviation. Among the most useful tools for realizing such sums are pulsed gradient fields.

Gradient Fields. Modern NMR spectrometers are equipped with the capability of applying a magnetic field gradient in any direction for a chosen, brief amount of time. If the direction is along the sample's z-axis, then while the gradient is on, the field varies as $B(z) = B_0 + \gamma z B_1$, where B_0 is the strong, external field and B_1 is the gradient power.

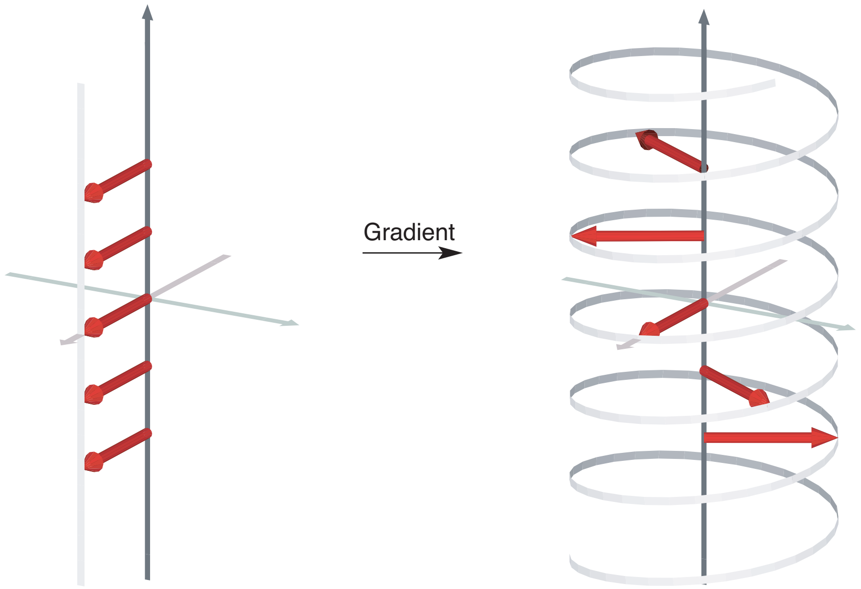


Figure 8. Pulsed Gradient Field along the *z*-Axis
Initial *x*-magnetization is assumed. A spin at $z = 0$ is not affected, but the ones above and below are rotated by an amount proportional to *z*. As a result, the local planar magnetization follows a spiral curve.

As a result of this gradient, the precession frequency of nuclear spins depends on their positions' *z*-coordinates. One of the most important applications of gradients is NMR imaging because gradients make it possible to distinguish different parts of the sample.

The effect of applying a *z*-gradient can be visualized for the situation in which there is only one observable nuclear spin per molecule. Suppose that the initial deviation density matrix of each nuclear spin is σ_x in the rotating frame. After a gradient pulse of duration t, the deviation of a nuclear spin at position z is given by $e^{-i\sigma_z vzt/2}\sigma_x e^{i\sigma_z vzt/2} = \cos(vzt)\sigma_x + \sin(vzt)\sigma_y$, where the constant v depends linearly on the strength of the gradient and the magnetic moment of the nucleus—see Figure 8. The effect of the gradient is a *z*-dependent change in phase. The coil used to measure planar magnetization integrates the contribution to the magnetization of all the nuclei in the neighborhood of the coil. Assuming a coil equally sensitive over the interval between $-a$ and a along the sample's *z*-axis, the observed total *x*-magnetization is

$$\begin{aligned} M_x &= \int_{-a}^{a} dz\ \mathrm{tr}\left(\sigma_x\left(\cos(vzt)\sigma_x + \sin(vzt)\sigma_y\right)\right) \\ &= \int_{-a}^{a} dz\ \mathrm{tr}\left(\cos(vzt)\sigma_x^2 + \sin(vzt)\sigma_x\sigma_y\right) \\ &= \int_{-a}^{a} dz\ \mathrm{tr}\left(\cos(vzt) + i\sin(vzt)\sigma_z\right) \\ &= 2\int_{-a}^{a} dz\ \cos(vzt)\ . \end{aligned} \tag{21}$$

For large values of vt, $M_x \cong 0$. In general, a sufficiently powerful gradient pulse eliminates the planar magnetization.

Interestingly, the effect of a a gradient pulse can be reversed if an opposite gradient pulse is applied for the same amount of time. This effect is called a "gradient echo." The reversal only works if the second pulse is applied sufficiently soon. Otherwise, diffusion randomizes the molecules' positions along the gradient's direction before the second pulse. If the positions are randomized, the phase change from the second pulse is no longer correlated with that from the first for any given molecule. The loss of memory of the phase change from a gradient pulse can be fine-tuned by variations in the delay between the two pulses in a gradient echo sequence. This method can be used for applying a controllable amount of phase noise, which is useful for investigating the effects of noise and the ability to correct for noise in QIP.

If the gradient pulse is not reversed and the memory of the phase changes is lost, then the pulse's effect can be described as an irreversible operation on the state of the nuclear spin. If the initial state of the nuclear spin in each molecule is ρ, then after the gradient pulse, the spin state of a molecule at position z is given by $\rho(z) = e^{-i\sigma_z vzt/2}\rho e^{i\sigma_z vzt/2}$. Suppose that the positions of the molecules are randomized over the region that the coil is sensitive to. Now it is no longer possible to tell where a given molecule was when the gradient pulse was applied. As a result, as far as our observations are concerned, the state of a molecule is given by $\rho(z)$, where z is random. In other words, the state is indistinguishable from

$$\rho' = \frac{1}{2a}\int_{-a}^{a} dz\rho(z) = \frac{1}{2a}\int_{-a}^{a} dz e^{-i\sigma_z vzt/2}\rho e^{i\sigma_z vzt/2} \quad . \tag{22}$$

Thus, the effect of the gradient pulse is equivalent to the operation $\rho \rightarrow \rho'$ as defined by the above equation. This is an operation of the type mentioned at the end of the previous section and can be used for making states such as pseudopure states. Note that, after the gradients have been turned off, nuclei at different positions cannot be distinguished by the measurement coil. It is therefore not necessary to wait for the molecules' positions to be randomized.

So far, we have described the effects of gradient pulses on isolated nuclear spins in a molecule. In order to restrict the effect to a single nuclear spin in a molecule, one can invert the other spins between a pair of identical gradient pulses in the same direction. This technique refocuses the gradient for the inverted spins. An example of how effects involving multiple nuclear spins can be exploited is the algorithm for pseudopure state preparation described in the section "Creating a Labeled Pseudopure State."

Examples of Quantum Algorithms for NMR

We give three examples of algorithms for NMR QIP. The first is an NMR implementation of the controlled-not gate. The second consists of a procedure for preparing a type of pseudopure state. And the last shows how NMR can be used to investigate the behavior of simple error-correction procedures. The first two examples are fundamental to QIP with NMR. Realizations of the controlled-not are needed to translate standard quantum algorithms into the language of NMR, and procedures for making pseudopure states have to precede the implementation of many quantum algorithms.

The Controlled-not. One of the standard gates used in quantum algorithms is the controlled-not. The controlled-not gate (**cnot**) acts on two qubits. The action of **cnot** can be described by "if the first qubit is $|1\rangle$, then flip the second qubit." Consequently, the effect of **cnot** on the logical states is given by the mapping

$$\begin{aligned}
\mathbf{cnot}\left|00\right\rangle &= \left|00\right\rangle \\
\mathbf{cnot}\left|01\right\rangle &= \left|01\right\rangle \\
\mathbf{cnot}\left|10\right\rangle &= \left|11\right\rangle \\
\mathbf{cnot}\left|11\right\rangle &= \left|10\right\rangle \quad .
\end{aligned} \tag{23}$$

As an operator, the controlled-not is given by

$$\textbf{cnot} = |0\rangle_1{}^1\langle 0| + |1\rangle_1{}^1\langle 1|\sigma_x^{(2)} = \left(\left(\mathbb{1} + \sigma_z^{(1)}\right) + \left(\mathbb{1} - \sigma_z^{(1)}\right)\sigma_x^{(2)}\right)\Big/2 \ . \tag{24}$$

The goal is to derive a sequence of NMR operations that realize the controlled-not. As discussed earlier ("Principles of Liquid-State NMR QIP"), the unitary operations implementable by simple NMR techniques are rotations $e^{-i\sigma_u{}^{(a)}\theta/2}$ by θ around the *u*-axis, where *u* is any direction in the plane (rf pulses), and the two-qubit operations $e^{-i\sigma_z{}^{(b)}\sigma_z{}^{(c)}\phi/2}$ (the *J*-coupling). We call $e^{-i\sigma_z{}^{(b)}\sigma_z{}^{(c)}\phi/2}$ a rotation by ϕ around $\sigma_z{}^{(b)}\sigma_z{}^{(c)}$. This terminology reflects the fact that such rotations and their effects on deviation density matrices can be understood by a generalization of the Bloch sphere picture called the product operator formalism introduced by Sørensen et al. (1983).

To implement the controlled-not using NMR techniques, one can decompose the gate into a sequence of 90° rotations around the main axes on each of the two qubits, and a 90° rotation around $\sigma_z{}^{(1)}\sigma_z{}^{(2)}$. One way to find a decomposition is to first realize that the two-qubit 90° rotation $e^{-i\sigma_z{}^{(1)}\sigma_z{}^{(2)}\pi/4}$ is equivalent to a combination of two gates, each conditional on the logical state of qubit 1. The first gate applies a 90° rotation around the *z*-axis ($e^{-i\sigma_z{}^{(2)}\pi/4}$) to qubit 2 conditional on qubit l's state being $|0\rangle_1$. The second applies the –90° rotation $e^{i\sigma_z{}^{(2)}\pi/4}$ to qubit 2 conditional on qubit l's state being $|1\rangle_1$. By following the two-qubit rotation with a –90° rotation around the *z*-axis ($e^{i\sigma_z{}^{(2)}\pi/4}$) on qubit 2, the total effect is to cancel the rotation if qubit 1 is in state $|0\rangle_1$; if qubit 1 is in state $|1\rangle_1$, the rotations add to a –180° rotation $e^{i\sigma_z{}^{(2)}\pi/2} = i\sigma_z{}^{(2)}$ on qubit 2. If we precede this sequence with $e^{-i\sigma_y{}^{(2)}\pi/4}$ and follow it by $e^{i\sigma_y{}^{(2)}\pi/4}$ (this operation is called conjugating by a –90° *y*-rotation), the overall effect is a conditional $-i\sigma_x{}^{(2)}$ operation. Note how the conjugation rotated the operation's axis according to the Bloch sphere rules. The controlled-not is obtained by eliminating the $-i$ with a 90° *z*-rotation on qubit 1. That is, the effect of the complete sequence is $e^{-i\pi/4}|0\rangle_1{}^1\langle 0| + e^{-i\pi/4}|1\rangle_2{}^2\langle 1|\sigma_x{}^{(2)}$, which is the controlled-not up to a global phase. The decomposition thus obtained can be represented as a quantum network with rotation gates, as shown in Figure 9. The corresponding NMR pulse sequence implementation is shown in Figure 10.

The effect of the NMR pulse sequence that implements the controlled-not can be visualized for logical initial states with the help of the Bloch-sphere representation of the states. Figure 11 shows such a visualization for two initial states.

The effects of the pulse sequence for the controlled-not can be shown with the Bloch sphere (Figure 11) only if the intermediate states are products of states on each qubit. Things are no longer so simple if the initial state of the spins is $1/\sqrt{2}(|0\rangle + |1\rangle)\,|0\rangle = 1/\sqrt{2}(|00\rangle + |10\rangle)$, for example. This is representable as spin 1's arrow pointing along the *x*-axis, but the *J*-coupling leads to a superposition of states (a maximally entangled state) no longer representable by a simple combination of arrows in the Bloch sphere.

Creating a Labeled Pseudopure State. One way to realize the standard pseudopure state starting from the equilibrium density matrix ρ_{thermal} is to eliminate the observable contributions due to terms of ρ_{thermal} different from $|0\ldots0\rangle\langle 0\ldots0|$. There are several different methods of accomplishing this task. For example, one can perform multiple experiments with different preprocessing of the equilibrium state so that signals from unwanted terms average to zero (temporal averaging), or one can use gradients to remove the unwanted terms in one experiment (spatial averaging).

In this section, we show how to use spatial averaging to prepare a so-called labeled pseudopure state on two nuclear spins. In general, instead of preparing the standard pseudopure state with deviation $|0\ldots\rangle\langle 0\ldots|$ on *n* spin-1/2 nuclei, one can prepare a

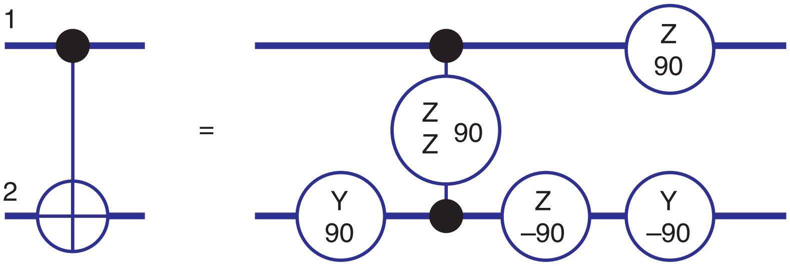


Figure 9. Quantum Network for Implementing the Controlled-not with NMR Operations

The conventions for depicting gates are as explained in the article "Quantum Information Processing" on page 2. The two one-qubit *z*-rotations can be implemented by a change in the reference phase of the rotating frame without any rf pulses being applied.

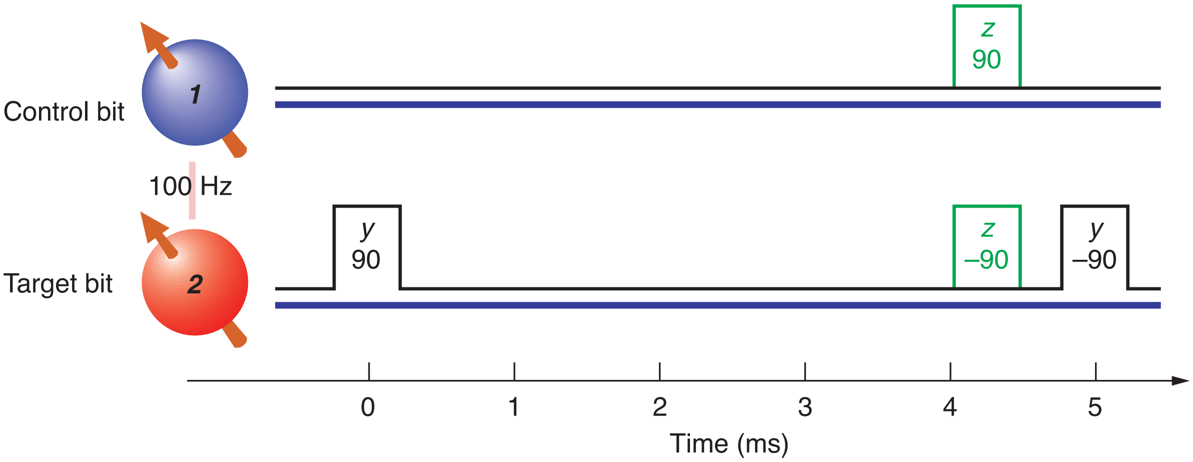


Figure 10. Pulse Sequence for Realizing the Controlled-not

The control bit is spin 1 and the target is spin 2. The pulses are shown with the representation introduced in Figure 6. The *z*-pulses (shown in green) are virtual, requiring only a change of reference frame. The placement of the *z*-pulses between the rf pulses is immaterial because they commute with the coupling that evolves in between. The delay between the two rf pulses is $1/(2J)$ (5 ms if $J = 100$ Hz), which realizes the desired two-qubit rotation by internal evolution. The –90° *y*-rotation is actually implemented with a 90° pulse with axis $-y$. The resulting rotation has the desired effect up to a global phase. The pulse widths are exaggerated and should be as short as possible to avoid errors due to coupling evolution during the rf pulses. Alternatively, techniques can be used that compensate for some of these errors (Knill et al. 2000).

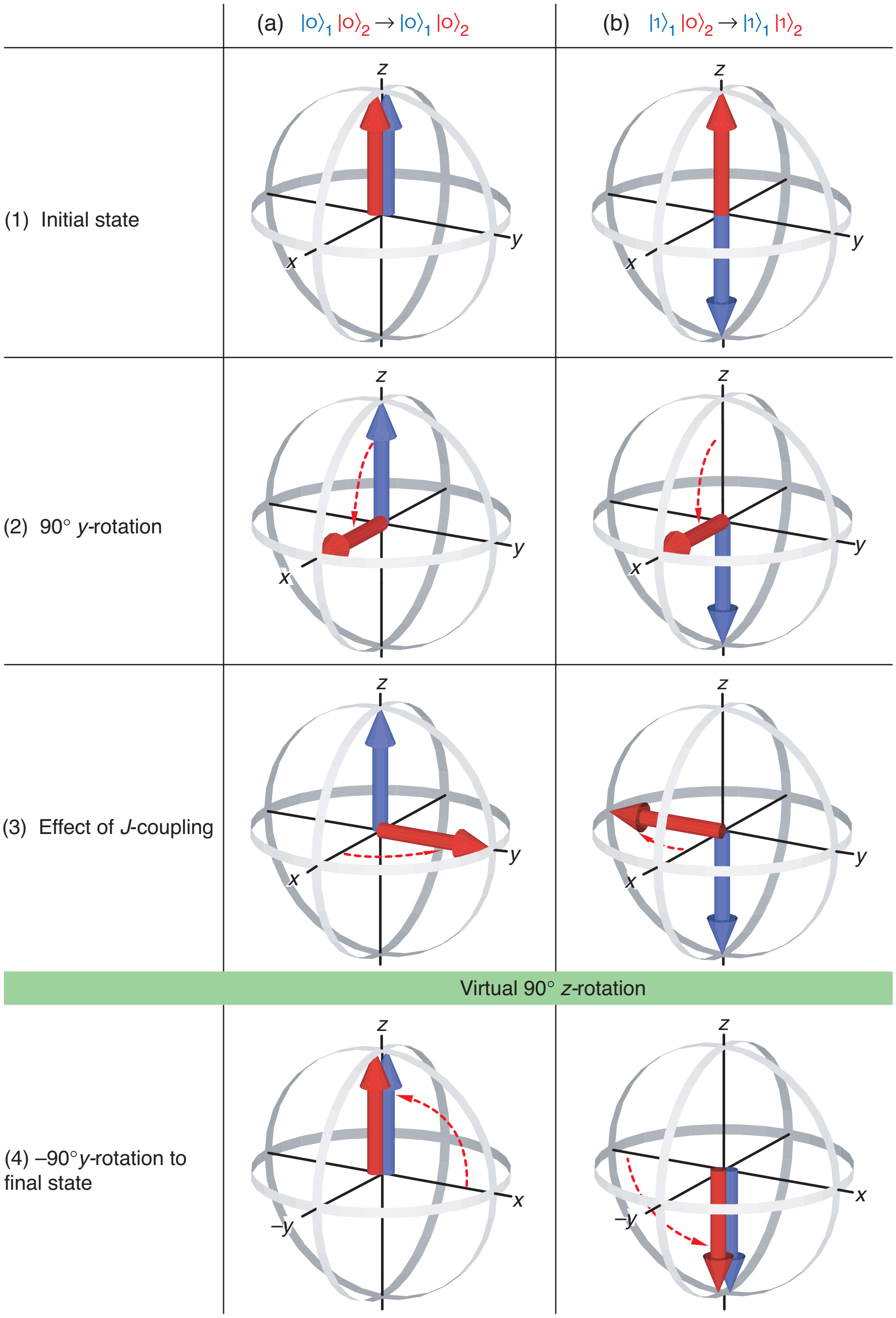

(a) $|0\rangle_1 |0\rangle_2 \rightarrow |0\rangle_1 |0\rangle_2$
(b) $|1\rangle_1 |0\rangle_2 \rightarrow |1\rangle_1 |1\rangle_2$
(1) Initial state
(2) 90° y-rotation
(3) Effect of J-coupling
Virtual 90° z-rotation
(4) –90° y-rotation to final state
z
x
y
–y

Figure 11. States Corresponding to the Controlled-not Pulse Sequence

The two columns (a) and (b) show the evolution of the qubit states during the controlled-not pulse sequence. The blue and red arrows represent spin 1 and 2, respectively. The configurations in rows 1 to 4 are shown (1) at the beginning of the sequence, (2) after the 90° *y*-rotation, (3) after the *J*-coupling (but before the *z*- and *y*-pulses), and (4) at the end of the sequence. The conditional effect is realized by the second spin's pointing down at the end of the second column. The effect of the *J*-coupling causing the evolution from 2 to 3 is best understood as a conditional rotation around the *z*-axis (forward by 90° if the first spin is up; backward, if it is down).

labeled pseudopure state with deviation $\sigma_x^{(1)}|0\ldots\rangle\langle 0\ldots|$ on $n + 1$ spins. This state is easily recognizable with an NMR observation of the first spin: Assuming that all the peaks arising from couplings to other spins are resolved, the first spin's peak group has 2^n peaks, corresponding to which logical states the other spins are in. If the current state is the labeled pseudopure state just mentioned, then all the other spins are in the logical state $|0\rangle$, which implies that, in the spectrum, only one of the peaks of the first spin's peak group is visible (see Figure 12).

The labeled pseudopure state can be used as a standard pseudopure state on n qubits. Observation of the final answer of a computation is possible by observing spin 1, provided that the coupling to the answer-containing spin is sufficiently strong for the peaks corresponding to its two logical states to be well separated. For this purpose, the couplings to the other spins need not be resolved in the peak group. Specifically, to determine the answer of a computation, the peaks of the spin 1 peak group are separated into two subgroups, the first (second) containing the peaks associated with the answer-containing spin being in state $|0\rangle$ ($|1\rangle$), respectively. Comparing the total signal in each of the two peak subgroups gives the relative probabilities of the two answers (0 or 1).

The labeled pseudopure state can also be used to investigate the effect of a process that manipulates the state of one qubit and requires n additional initialized qubits. Examples include experimental verification of one-qubit error-correcting codes as explained in the next section.

For preparing the two-qubit labeled pseudopure state, consider the two carbon nuclei in labeled TCE with the proton spin decoupled so that its effect can be ignored. A "transition" in the density matrix for this system is an element of the density matrix of the form $|ab\rangle\langle cd|$, where a, b, c, and d are 0 or 1. Let $\Delta(ab,\ cd) = (a - c) + (b - d)$, where in the expression on the right, a, b, c, and d are interpreted as the numbers 0 or 1, as appropriate. Applying a pulsed gradient along the z-axis evolves the transitions according to $|ab\rangle\langle cd| \rightarrow e^{i\Delta(ab,cd)\nu z}|ab\rangle\langle cd|$, where ν is proportional to the product of the gradient power and pulse time and z is the molecule's position along the z-coordinate. For example, $|01\rangle\langle 10|$ has $\Delta = 0$ and is not affected whereas $|00\rangle\langle 11|$ acquires a phase of $e^{-i2\nu z}$. There are only two transitions, $|00\rangle\langle 11|$ and $|11\rangle\langle 00|$, whose acquired phase has a rate of $\Delta = \pm 2$ along the z-axis. These transitions are called 2-coherences because $\Delta = \pm 2$. The idea is to first recognize that these transitions can be used to define a labeled pseudopure "cat" state (see below), then to exploit the 2-coherences' unique behavior under the gradient in order to extract the pseudopure cat state, and finally to decode to a standard labeled pseudopure state. Note that the property that 2-coherences' phases evolve at twice the basic rate is a uniquely quantum phenomenon for two spins. No such effect is observed for a pair of classical spins.

The standard two-qubit labeled pseudopure state's deviation can be written as $\rho_{\mathrm{std}x} = \sigma_x^{(1)}1/2(\mathbb{1} + \sigma_z^{(2)})$. We can consider other deviations of this form where the two Pauli operators are replaced by a pair of different commuting products of Pauli operators. An example is

$$\rho_{\mathrm{cat}_x} = \left(\sigma_x^{(1)}\sigma_x^{(2)}\right)\frac{1}{2}\left(\mathbb{1} + \sigma_z^{(1)}\sigma_z^{(2)}\right) , \tag{25}$$

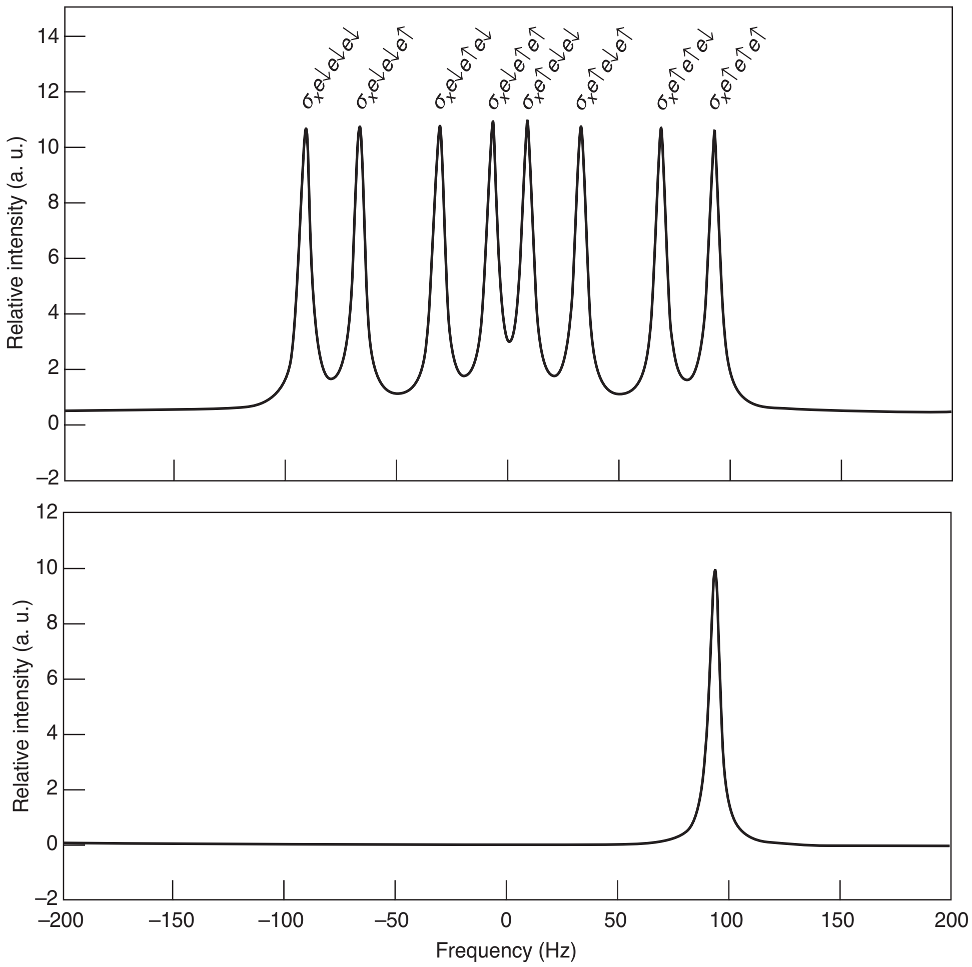


Figure 12. Labeled Pseudopure State Spectrum vs Peak Group
(a) This spectrum shows the peak group of a simulated nuclear spin coupled to three other spins with coupling constants of 100 Hz, 60 Hz, and 24 Hz. The simulation parameters are the same as in Figure 7. Given above each peak is the part of the initial deviation that contributes to the peak. The spin labels have been omitted. Each contributing deviation consists of σ_x on the observed nucleus followed by one of the logical (up or down) states (density matrices) for each of the other spins. The notation is as defined after Equation 9.
(b) This spectrum shows what is observed if the initial deviation is the standard labeled pseudopure state. This state contributes only to the rightmost peak, as this peak is associated with the logical $|0\rangle$ states on the spins not observed.

where we replaced $\sigma_x^{(1)}$ by $\sigma_x^{(1)}\sigma_x^{(2)}$ and $\sigma_z^{(2)}$ by $\sigma_z^{(1)}\sigma_z^{(2)}$. As announced, the two Pauli products commute. We will show that there is a simple sequence of 90° rotations whose effect is to decode the deviations $\sigma_x^{(1)}\sigma_x^{(2)} \rightarrow \sigma_x^{(1)}$ and $\sigma_z^{(1)}\sigma_z^{(2)} \rightarrow \sigma_z^{(2)}$, thus converting the state ρ_{cat_x} to ρ_{std_x}. The state ρ_{cat_x} can be expressed in terms of the transitions as follows:

$$\rho_{\text{cat}_x} = |00\rangle\langle 11| + |11\rangle\langle 00| \; . \qquad (26)$$

It can be seen that ρ_{cat_x} consists only of 2-coherences. Another such state is

$$\rho_{\text{cat}_y} = \left(\sigma_x^{(1)}\sigma_y^{(2)}\right)\frac{1}{2}\left(\mathbb{1} + \sigma_z^{(1)}\sigma_z^{(2)}\right) \qquad (27)$$

$$= -i|00\rangle\langle 11| + i|11\rangle\langle 00| \; . \qquad (28)$$

Suppose that one can create a state that has a deviation of the form $\rho = \alpha\rho_{\mathrm{cat}x} + \beta\rho_{\mathrm{rest}}$ such that ρ_{rest} contains no 2-coherences or 0-coherences. After a gradient pulse is applied, the state becomes

$$\alpha\Big(\cos(2\nu z)\,\rho_{\mathrm{cat}_x} + \sin(2\nu z)\,\rho_{\mathrm{cat}_y}\Big) + \beta\,\rho_{\mathrm{rest}}(z)\ , \tag{29}$$

where $\rho_{\mathrm{rest}}(z)$ depends periodically on z with spatial frequencies of $\pm\nu$, not $\pm 2\nu$ or 0. We can then decode this state to

$$\rho \;=\; \alpha\Big(\cos(2\nu z)\,\rho_{\mathrm{std}_x} + \sin(2\nu z)\,\rho_{\mathrm{std}_y}\Big) + \beta\rho'_{\mathrm{rest}}(z) \tag{30}$$

$$\;=\; \alpha\Big(\cos(2\nu z)\,\sigma_x^{(1)} + \sin(2\nu z)\,\sigma_y^{(1)}\Big)\frac{1}{2}\Big(\mathbb{1} + \sigma_z^{(1)}\Big) + \beta\rho'_{\mathrm{rest}}(z)\ . \tag{31}$$

If one now applies a gradient pulse of twice the total strength and opposite orientation, the first term is restored to $\alpha\rho_{\mathrm{std}x}$, but the second term retains nonzero periodicities along z. Thus, if we no longer use any operations to distinguish among different molecules along the z-axis or if we let diffusion erase the memory of the position along z, then the second term is eliminated from observability by being averaged to 0. The desired labeled pseudopure state is obtained. Zero-coherences during the initial gradient pulse are acceptable provided that the decoding transfers them to coherences different from 0 or 2 during the final pulse in order to ensure that they also average to 0. A pulse sequence that realizes a version of the above procedure is shown in Figure 13.

We can follow what happens to an initial deviation density matrix of $\sigma_z^{(1)}$ as the network of Figure 13 is executed. We use product operators with the abbreviations $I = \mathbb{1}$, $X = \sigma_x$, $Y = \sigma_y$, $Z = \sigma_z$ and, for example, $XY = \sigma_x^{(1)}\sigma_y^{(2)}$. At the checkpoints indicated in the figure, the deviations are the following:

Checkpoints

$$\begin{array}{rlll}
1 & ZI & & \\
2 & XI & & \\
3 & YZ & & \\
4 & YX \propto & & \\
 & YX + XY & + & YX - XY \\
5 & \cos(2\nu z)(YX + XY) + \sin(2\nu z)(YY - XX) & + & YX - XY \\
6 & \cos(2\nu z)(YZ + XY) + \sin(2\nu z)(YY - XZ) & + & YZ - XY \\
7 & \cos(2\nu z)(-XI + XY) + \sin(2\nu z)(YY - YI) & + & -XI - XY \\
8 & \cos(2\nu z)(-XI - XZ) + \sin(2\nu z)(-YZ - YI) & + & -XI + XZ \\
9 & -X(I + Z) & + & -\big(\cos(-2\nu z)X + \sin(-2\nu z)Y\big)\big(I - Z\big)\ .
\end{array} \tag{32}$$

Except for a sign, the desired state is obtained. The rightmost term is eliminated after integrating over the sample or after diffusion erases memory of z.

This method for making a two-qubit labeled pseudopure state can be extended to arbitrarily many (n) qubits by exploiting the two n-coherences, which are the transitions with $\Delta = \pm n$. An experiment implementing this method can be used to determine how good the available quantum control is. The quality of the control is determined by a comparison of two spectral signals: I_p, the intensity of the single peak that shows up in the peak group for spin 1 when observing the labeled pseudopure state, and I_0, the intensity of the same peak in an observation of the initial deviation after applying a 90° pulse

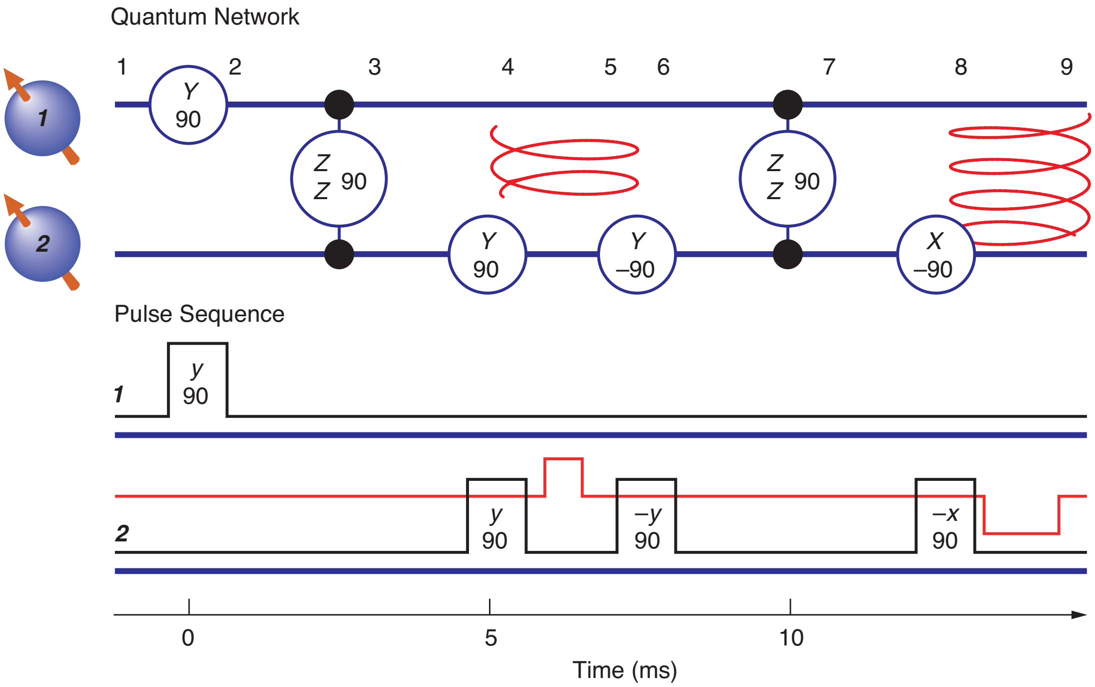


Figure 13. Realizing a Two-Qubit Labeled Pseudopure State

The network is shown above the pulse sequence realizing it. A coupling constant of 100 Hz is assumed. Gradients are indicated by spirals in the network. The gradient strength is given as the red line in the pulse sequence. The doubling of the integrated gradient strength required to achieve the desired "echo" is indicated by a doubling of the gradient pulse time. The numbers above the quantum network are checkpoints used in the discussion below. The input state's deviation is assumed to be $\sigma_z^{(1)}$. This deviation can be obtained from the equilibrium state by applying a 90° rotation to spin 2 followed by a gradient pulse along another axis to remove $\sigma_z^{(2)}$. Instead of using a gradient pulse, one can use phase cycling, which involves performing two experiments, the second having the sign of the phase in the first *y*-pulse changed, and then subtracting the measured signals.

to rotate $\sigma_z^{(1)}$ into the plane. We performed this experiment on a seven-spin system and determined that $I_p/I_0 = .73 \pm .02$. This result implies a total error of 27 ± 2 percent. Because the implementation has 12 two-qubit gates, an error rate of about 2 percent per two-qubit gate is achievable for nuclear spins in this setting (Knill et al. 2000).

Quantum Error Correction for Phase Errors. Currently envisaged scalable quantum computers require the use of quantum error correction to enable relatively error-free computation on a platform of physical systems that are inherently error prone. For this reason, some of the most commonly used subroutines in quantum computers will be associated with maintaining information in encoded forms. This observation motivates experimental realizations of quantum error correction to determine whether adequate control can be achieved in order to implement these subroutines and to see in a practical setting that error correction has the desired effects. Experiments to date have included realizations of a version of the three-qubit repetition code (Cory et al. 1998) and of the five-qubit one-error-correcting code (the shortest possible such code)—see the article "Quantum Information Processing" on page 2. In this section, we discuss the experimental implementation of the former.

In NMR, one of the primary sources of error is phase decoherence of the nuclear spins due to both systematic and random fluctuations in the field along the *z*-axis. At the same time, using gradient pulses and diffusion, phase decoherence is readily induced artificially and in a controlled way. The three-bit quantum repetition code (see the article "Introduction to Quantum Error Correction" on page 188) can be adapted to protect against phase errors to first order. Define $|+\rangle = 1/\sqrt{2}\,(|0\rangle + |1\rangle)$ and $|-\rangle = 1/\sqrt{2}\,(|0\rangle - |1\rangle)$. The code we want is defined by the logical states

$$|0\rangle_L = |+\rangle|+\rangle|+\rangle \ , \text{ and } |1\rangle_L = |-\rangle|-\rangle|-\rangle \ . \tag{33}$$

It is readily seen that the three one-qubit phase errors $\sigma_z^{(1)}$, $\sigma_z^{(2)}$, and $\sigma_z^{(3)}$ and "no error" ($\mathbb{1}$) unitarily map the code to orthogonal subspaces. It follows that this set of

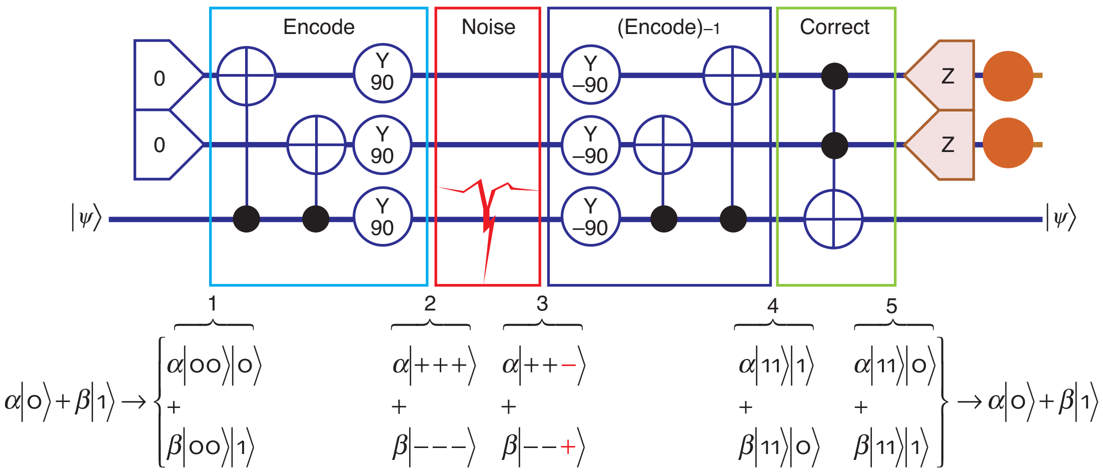


Figure 14. Quantum Network for the Three-Qubit Phase-Error-Correcting Repetition Code
The bottom qubit is encoded with two controlled-nots and three *y*-rotations. In the experiment, either physical or controlled noise is allowed to act. The encoded information is then decoded. For the present purposes, it is convenient to separate the decoding procedures into two steps: The first is the inverse of the encoding procedure; the second consists of a Toffoli gate that uses the error information in the syndrome qubits (the top two) to restore the encoded information. The Toffoli gate in the last step flips the output qubit conditionally on the syndrome qubits' state being |11⟩. This gate can be realized with NMR pulses and delays by using more sophisticated versions of the implementation of the controlled-not. The syndrome qubits can be "dumped" at the end of the procedure. The behavior of the network is shown for a generic state in which the bottom qubit experiences a σ_z error.

errors is correctable (for a full discussion, see the article "Introduction to Error Correction" on page 188). The simplest way to use this code is to encode one qubit's state into it, wait for some errors to happen, and then decode to an output qubit. Success is indicated by the output qubit's state being significantly closer to the input qubit's state after error correction. Without errors between encoding and decoding, the output state should be the same as the input state, provided that the encoding and decoding procedures are implemented perfectly. Therefore, in this case, the experimentally determined difference between input and output gives a measurement of how well the procedures were implemented.

To obtain the phase-correcting repetition code from the standard repetition code, we apply Hadamard transforms or 90° *y*-rotations to each qubit. The quantum network shown in Figure 14 was obtained in this fashion from the network given in the article on error correction.

To determine the behavior and the quality of the implementation for various σ_z-error models in an actual NMR realization, one can use as initial states labeled pseudopure states with deviations $\sigma_u|00\rangle\langle 00|$ for $u = x, y, z$. Without error, the total output signal on spin 1 along σ_u for each u should be the same as the input signal. Some of the data reported by Cory and coworkers (1998) are shown in Figure 15.

Work on benchmarking error-control methods using liquid-state NMR is continuing. Other experiments include the implementation of a two-qubit code with an application to phase errors (Leung et al. 1999) and the verification of the shortest nontrivial noiseless subsystem on three qubits (Viola et al. 2001). The latter demonstrates that, for some physically realistic noise models, it is possible to store quantum information in such a way that it is completely unaffected by the noise.

Discussion

Overview of Contributions to QIP. Important issues in current experimental efforts toward realizing QIP are to find ways of achieving necessary quantum control and to determine whether sufficiently low error rates are possible. Liquid-state NMR is the only extant system (as of 2002) with the ability to realize relatively universal manipulations on more than two qubits—restricted control has been demonstrated in four ions (Sackett et al 2000). For this reason, NMR serves as a useful platform for developing and experimentally verifying techniques for QIP and for establishing simple procedures for benchmarking information-processing tasks. The cat state and the various error-

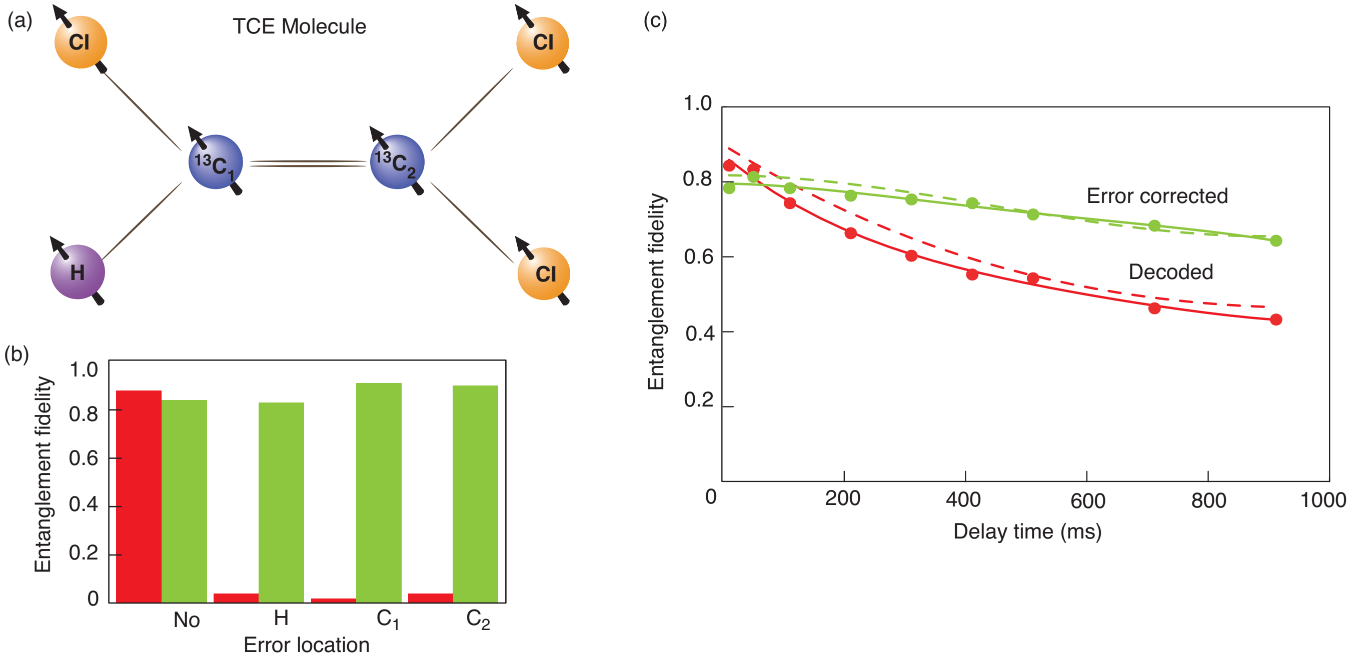


Figure 15. Experimental Fidelities of Error Correction

(a) The molecule used in the experiment is shown here. (b) The bar graph shows fidelities for explicitly applied errors. The fidelities f (technically, the entanglement fidelities) are an average of the signed ratios f_u of the input to the output signals for the initial deviations $\sigma_u|oo\rangle\langle oo|$ with $u = x, y, z$. Specifically, $f = 1/4(1 + f_x + f_y + f_z)$. The reduction from 1 of the green bars (showing fidelity for the full procedure) is due to errors in our implementation of the pulses and from relaxation processes. The red bars are the fidelity for the output before the last error-correction step, and they contain the effects of the errors. (c) The graph shows the fidelities for the physical relaxation process. Here, the evolution consisted of a delay of up to 1000 ms. The red curve is the fidelity of the output qubit before the final Toffoli gate that corrects the errors based on the syndrome. The green curve is the fidelity of the output after the Toffoli gate. The effect of error correction can be seen by a significant flattening of the curve because correction of first-order (that is, single) phase errors implies that residual, uncorrected (double or triple) phase errors increase quadratically in time. The green curve starts lower than the red one because of additional errors incurred by the implementation of the Toffoli gate. The dashed curves are obtained by simulation using estimated phase relaxation rates with halftimes of 2 s (proton), 0.76 s (first carbon) and 0.42 s (second carbon). Errors in the data points are approximately 0.05. (For a more thorough implementation and analysis of a three-qubit phase-error-correcting code, see Sharf et al. 2000).

correction benchmarks (Knill et al. 2000, Knill et al. 2001) consist of a set of quantum control steps and measurement procedures that can be used with any general-purpose QIP system to determine, in a device-independent way, the degree of control achieved. The demonstration of error rates in the few percent per nontrivial operation is encouraging. For existing and proposed experimental systems other than NMR, achieving such error rates is still a great challenge.

Prior research in NMR, independent of quantum information, has proved to be a rich source of basic quantum-control techniques useful for physically realizing quantum information in other settings. We mention four examples. The first is the development of sophisticated shaped-pulse techniques that can selectively control transitions or spins while being robust against typical errors. These techniques are finding applications to quantum control involving laser pulses (Warren et al. 1993) and are likely to be very useful when using coherent light to accurately control transitions in atoms or quantum dots, for example. The second is the recognition that there are simple ways in which imperfect pulses can be combined to eliminate systematic errors such as those associated with miscalibration of power or side effects on off-resonant nuclear spins. Although many of these techniques were originally developed for such problems as accurate inversion of spins, they are readily generalized to other quantum gates (Levitt 1982, Cummins and Jones 1999). The third example is decoupling used to reduce unwanted external interactions. For example, a common problem in NMR is to eliminate the interactions between proton and labeled carbon nuclear spins in order to observe decoupled carbon spins. In this case, the protons constitute an external system with an unwanted interaction. To eliminate the interaction, it is sufficient to invert the protons frequently. Sophisticated techniques for ensuring that the interactions are effectively turned off independent of pulse errors have been developed (Ernst et al. 1994). These techniques have been greatly generalized and shown to be useful for actively creating protected qubit subsystems in any situation in which the interaction has relatively long correlation times (Viola and Lloyd 1998, Viola et al. 1999). Refocusing to undo unwanted internal interactions is our fourth example. The technique for turning off the coupling between spins that is so important for realizing QIP in liquid-state NMR is a special case of much more general methods of turning off or refocusing Hamiltonians. For example, a famous technique in solid-state NMR is to reverse the dipolar coupling Hamiltonian using a clever sequence of 180° pulses at different phases (Ernst et al. 1994, page 48). Many other proposed QIP systems suffer from such internal interactions while having similar control opportunities.

The contributions of NMR QIP research extend beyond those directly applicable to experimental QIP systems. It is due to NMR that the idea of ensemble quantum computation with weak measurement was introduced and recognized as being, for true pure initial states, as powerful for solving algorithmic problems as the standard model of quantum computation. (It cannot be used in settings involving quantum communication.) One implication is that, to a large extent, the usual assumption of projective measurement can be replaced by any measurement that can statistically distinguish between the two states of a qubit. Scalability still requires the ability to reset qubits during the computation, which is not possible in liquid-state NMR. Another interesting concept emerging from NMR QIP is that of computational cooling (Schulman and Vazirani 1998), which can be used to efficiently extract initialized qubits from a large number of noisy qubits in initial states that are only partially biased toward $|0\rangle$. This is a very useful tool for better exploiting otherwise noisy physical systems.

The last example of interesting ideas arising from NMR studies is the one-qubit model of quantum computation (Knill and Laflamme 1998). This is a useful abstraction of the capabilities of liquid-state NMR. In this model, it is assumed that initially, one qubit is in the state $|0\rangle$ and all the others are in random states. Standard unitary quantum gates can be applied, and the final measurement is destructive. Without loss of generality, one can assume that all qubits are reinitialized after the measurement. This model can perform interesting physics simulations with no known efficient classical algorithms. On the other hand, with respect to oracles, it is strictly weaker than quantum computation. It is also known that it cannot faithfully simulate quantum computers (Ambainis et al. 2000).

Capabilities of Liquid-State NMR. One of the main issues in liquid-state NMR QIP is the highly mixed initial state. The methods for extracting pseudopure states are not practical for more than 10 (or so) nuclear spins. The problem is that for these methods, the pseudopure state signal decreases exponentially with the number of qubits prepared while the noise level is constant. This exponential loss limits the ability to explore and benchmark standard quantum algorithms even in the absence of noise. There are in fact ways in which liquid-state NMR can be usefully applied to many more qubits. The first and less practical is to use computational cooling for a (unrealistically) large number of spins to obtain less mixed initial states. Versions of this technique have been studied and used in NMR to increase signal to noise (Glaser et al. 1998). The second is to use the one-qubit model of quantum computation instead of trying to realize pseudopure states. For this purpose, liquid-state NMR is limited only by relaxation noise and pulse control errors, not by the number of qubits. Noise still limits the number of useful operations, but nontrivial physics simulations are believed to be possible with less than 100 qubits (Lloyd 1996). Remarkably, a one-qubit quantum computer can efficiently obtain a significant amount of information about the spectrum of a Hamiltonian that can be emulated on a quantum computer (Knill and Laflamme 1998, Somma et al. 2002, Miquel et al. 2002). Consequently, although QIP with molecules in liquid state cannot realistically be used to implement standard quantum algorithms involving more than about 10 qubits, its capabilities have the potential of exceeding the resource limitations of available classical computers for some applications.

Prospects for NMR QIP. There are many more algorithms and benchmarks that can be usefully explored using the liquid state NMR platform. We hope to soon have a molecule with ten or more useful spins and good properties for QIP. Initially, this molecule can be used to extend and verify the behavior of existing scalable benchmarks. Later, experiments testing basic ideas in physics simulation or more sophisticated noise-control methods are likely.

Liquid-state NMR QIP is one of many ways in which NMR can be used for quantum information. One of the promising proposals for quantum computation is based on phosphorus embedded in silicon (Kane 1998) and involves controlling phosphorus nuclear spins using NMR methods. In this proposal, couplings and frequencies are controlled with locally applied voltages. Universal control can be implemented with rf pulses. It is also possible to scale up NMR QIP without leaving the basic paradigms of liquid-state NMR while adding such features as high polarization, the ability to dynamically reset qubits (required for scalability), and much faster two-qubit gates. One proposal for achieving this goal is to use dilute molecules in a solid-state matrix instead of molecules in liquid (Cory et al. 2000). This approach may lead to pure-state quantum computation for significantly more than ten qubits.

NMR QIP has been a useful tool for furthering our understanding of the experimental challenges of quantum computation. We believe that NMR QIP will continue to shed light on important issues in physically realizing quantum information. ■

Raymond Laflamme graduated from the University of Laval, Canada, in 1983 with a Bachelor's degree in physics. Raymond received a Ph.D. in applied mathematics and theoretical physics from the University of Cambridge, England, where he worked under the direction of Stephen Hawking. Raymond then became a Killam postdoctoral fellow at the University of British Columbia before returning to Cambridge as a research fellow at Peterhouse College. In 1992, he came to Los Alamos as a Director-funded postdoctoral fellow, became an Oppenheimer fellow in 1994, and a technical staff member in 1997. In the fall of 2001, he moved to the newly created Perimeter Institute at the University of Waterloo, where he currently holds a Canadian Research Chair in Quantum Information. His research includes both theoretical and experimental investigations of quantum information processing.

David Cory received both his undergraduate and graduate degrees in chemistry from Case Western Reserve University. He is currently the Rasmussen Professor of Nuclear Engineering at the Massachusetts Institute of Technology. David and his students work to advance the instrumentation, methodology, and applications of nuclear magnetic resonance spectroscopy.

Further Reading

Ambainis, A., L. J. Schulman, and U. Vazirani. 2000. Computing with Highly Mixed States. In *Proceedings of the 32th Annual ACM Symposium on the Theory of Computation* (*STOC*). p. 697. New York:. ACM Press.

Anderson, A. G., and E. L. Hahn. 1955. Spin Echo Storage Technique. U.S. Patent # 2,714,714.

Anderson, A. G., R. Garwin, E. L. Hahn, J. W. Horton, and G. L. Tucker. 1955. Spin Echo Serial Storage Memory. *J. Appl. Phys*. **26**: 1324.

Barenco, A., C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, et al. 1995. Elementary Gates for Quantum Computation. *Phys. Rev. A* **52**: 3457.

Bloch, F. 1946. Nuclear Induction. *Phys. Rev*. **70**: 460.

Chuang, I. L., L. M. K. Vandersypen, X. Zhou, D. W. Leung, and S. Lloyd. 1998. Experimental Realization of a Quantum Algorithm. *Nature* **393**: 143.

Cory, D. G., A. F. Fahmy, and T. F. Havel. 1997. Ensemble Quantum Computing by NMR-Spectroscopy. *Proc. Natl. Acad. Sci. U.S.A*. **94**: 1634.

Cory, D. G., R. Laflamme, E. Knill, L. Viola, T. F. Havel, N. Boulant, et al. 2000. NMR Based Quantum Information Processing: Achievements and Prospects. *Fortschr. Phys*. **48**: 875.

Cory, D. G., W. Maas, M. Price, E. Knill, R. Laflamme, W. H. Zurek, et al. 1998. Experimental Quantum Error Correction. *Phys. Rev. Lett*. **81**: 2152.

Cummins, H. K., and J. A. Jones. 1999. Use of Composite Rotations to Correct Systematic Errors in NMR Quantum Computation. [Online]: http://eprints.lanl.gov (quant-ph/9911072).

DiVincenzo, D. P. 1995. Two-Bit Gates are Universal for Quantum Computation. *Phys. Rev. A* **51**: 1015.

Ernst, R. R., G. Bodenhausen, and A. Wokaun. 1994. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford: Oxford University Press.

Gershenfeld, N. A., and I. L. Chuang. 1997. Bulk Spin Resonance Quantum Computation. *Science* **275**: 350.

Glaser, S. J., T. Schulte-Herbruggen, M. Sieveking, 0. Schedletzky, N. C. Nielsen, 0. W. Sørensen, and C. Griesigner. 1998. Unitary Control in Quantum Ensembles: Maximizing Signal Intensity in Coherent Spectroscopy. *Science* **280**: 421.

Hahn, E. L. 1950. Spin Echoes. *Phys. Rev*. **80**: 580.

Jones, J. A., and E. Knill. 1999. Efficient Refocussing of One Spin and Two Spin Interactions for NMR. *J. Magn. Reson*. **141**: 322.

Jones, J. A., M. Mosca, and R. H. Hansen. 1998. Implementation of a Quantum Search Algorithm on a Quantum Computer. *Nature* **392**: 344.

Kane, B. E. 1998. A Silicon-Based Nuclear Spin Quantum Computer. *Nature* **393**: 133.

Knill, E., and R. Laflamme. 1998. On the Power of One Bit of Quantum Information. *Phys. Rev. Lett*. **81**: 5672.

Knill, E., I. L. Chuang, and R. Laflamme. 1998. Effective Pure States for Bulk Quantum Computation. *Phys. Rev. A* **57**: 3348.

Knill, E., R. Laflamme, R. Martinez, and C. Negrevergne. 2001. Implementation of the Five Qubit Error Correction Benchmark. *Phys. Rev. Lett*. **86**: 5811.

Knill, E., R. Laflamme, R. Martinez, and C.-H. Tseng. 2000. An Algorithmic Benchmark for Quantum Information Processing. *Nature* **404**: 368.

Leung, D. W., I. L. Chuang, F. Yamaguchi, and Y. Yamamoto. 1999. Efficient Implementation of Selective Recoupling in Heteronuclear Spin Systems Using Hadamard Matrices. [Online]: http://eprints.lanl.go. (quant-ph/9904100).

Leung, D., L. Vandersypen, X. L. Zhou, M. Sherwood, C. Yannoni, M. Kubinec, and I. L. Chuang. 1999. Experimental Realization of a Two-Bit Phase Damping Quantum Code. *Phys. Rev. A* **60**: 1924.

Levitt, M. H. 1982. Symmetrical Composite Pulse Sequences for NMR Population-Inversion 1. Compensation for Radiofrequency Field Inhomogeneity. *J. Magn. Reson*. **48**: 234.

Lloyd, S. 1995. Almost Any Quantum Logic Gate is Universal. *Phys. Rev. Lett*. **75**: 346.

Lloyd, S. 1996. Universal Quantum Simulators. *Science* **273**: 1073.

Mansfield, P., and P. Morris. 1982. NMR Imaging in Biomedicine. *Adv. Magn. Reson*. **S2**: 1.

Miquel, C., J. P. Paz, M. Saraceno, B. Knill, R. Laflamme, and C. Negrevergne. 2002. Interpretation of Tomography and Spectroscopy as Dual Forms of Quantum Computations. *Nature* **418**: 59.

Purcell, E. M., H. C. Torrey, and R. V. Pound. 1946. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev*. **69**: 37.

Sackett, C. A., D. Kielpinski, B. B. King, C. Langer, V. Meyer, C. J. Myatt, et al. 2000. Experimental Entanglement of Four Particles. *Nature* **404**: 256.

Schulman, L. J., and U. Vazirani. 1998. Scalable NMR Quantum Computation. In *Proceedings of the 3lth Annual ACM Symposium on the Theory of Computation* (*STOC*). p. 322. El Paso, TX: ACM Press.

Sharf, Y., T. F. Havel, and D. G. Cory. 2000. Spatially Encoded Pseudopure States for NMR Quantum-Information Processing. *Phys. Rev. A* **62**: 052314.

Sharf, Y., D. G. Cory, S. S. Somaroo, E. Knill, R. Laflamme, W. H. Zurek, and T. F. Havel. 2000. A Study of Quantum Error Correction by Geometric Algebra and Liquid-State NMR Spectroscopy. *Mol. Phys*. **98**: 1347.

Somma, R., G. Ortiz, J. E. Gubernatis, R. Laflamme, and E. Knill. 2002. Simulating Physical Phenomena by Quantum Networks. *Phys. Rev. A*. **65**:042323

Sørensen, O. W., G. W. Eich, M. H. Levitt, G. Bodenhausen, and R. R. Ernst. 1983. Product Operator-Formalism for the Description of NMR Pulse Experiments. *Prog. Nucl. Magn. Reson. Spectrosc*. **16**: 163.

Stoll, M. E., A. J. Vega, and R. W. Vaughan. 1977. Explicit Demonstration of Spinor Character for a Spin-1/2 Nucleus Using NMR Interferometry. *Phys. Rev. A* **16**: 1521.

Viola, L., and S. Lloyd. 1998. Dynamical Suppression of Decoherence in Two-State Quantum Systems. *Phys. Rev. A* **58**: 2733.

Viola, L., E. Knill, and S. Lloyd. 1999. Dynamical Decoupling of Open Quantum Systems. *Phys. Rev. Lett*. **82**: 2417.

Viola, L., E. M. Fortunato, M. A. Pravia, E. Knill, R. Laflamme, and D. G. Cory. 2001. Experimental Realization of Noiseless Subsystems for Quantum Information Processing. *Science* **293**: 2059.

Warren, W. S., H. Rabitz, and M. Dahleh. 1993. Coherent Control of Quantum Dynamics: The Dream is Alive. *Science* **259**: 1581.

Zalta, E. N., ed. 2002. *Stanford Encyclopedia of Philosophy*. Stanford University, Stanford, CA: The Metaphysics Research Lab.

Contact Information

R. Laflamme: laflamme@iqc.ca

E. Knill: knill@1anl.gov

D. Cory: dcory@mit.edu

E. M.Fortunato: evanmf@mit.edu

T. Havel: efhave1@mit.edu

C. Miquel: miquel@df.uba.ar

R. Martinez: rudy@lanl.gov

C. Negrevergne: cjn@lanl.gov

G. Ortiz: g_ortiz@lanl.gov

M. A.Pravia: praviam@mit.edu

Y. Sharf: ysharf@mit.edu

S. Sinha: suddha@mit.edu

R. Somma: somma@lanl.gov

L. Viola: lviola@lanl.gov

Glossary

Bloch sphere. A representation of the state space of a qubit using the unit sphere in three dimensions. See Figure 3 in the main text of the article.

Crosstalk. In using physical control to implement a gate, crosstalk refers to unintended effects on qubits not involved in the gate.

Decoupling. A method for turning off the interactions between two sets of spins. In NMR, this task can be achieved if one applies a rapid sequence of refocusing pulses to one set of spins. The other set of spins can then be controlled and observed as if independent of the first set.

Deviation of a state. If ρ is a density matrix for a state and $\rho = \alpha\mathbb{1} + \beta\sigma$, then σ is a deviation of ρ.

Ensemble computation. Computation with a large ensemble of identical and independent computers. Each step of the computation is applied identically to the computers. At the end of the computation, the answer is determined from a noisy measurement of the fraction p_1 of the computers whose answer is "1." The amount of noise is important for resource accounting: To reduce the noise to below ε requires increasing the resources used by a factor of the order of $1/\varepsilon^2$.

Equilibrium state. The state of a quantum system in equilibrium with its environment. In the present context, the environment behaves like a heat bath at temperature T, and the equilibrium state can be written as $\rho = e^{-H/kT}/Z$, where H is the effective internal Hamiltonian of the system and Z is determined by the identity $\mathrm{tr}\rho = 1$.

FID. Free induction decay. To obtain a spectrum on an NMR spectrometer after having applied pulses to a sample, one measures the decaying planar magnetization induced by the nuclear spins as they precess. The x- and y-components $M_x(t)$ and $M_y(t)$ of the magnetization as a function of time are combined to form a complex signal $M(t) = M_x(t) + iM_y(t)$. The record of $M(t)$ over time is called the FID, which is Fourier-transformed to yield the spectrum.

Inversion. A pulse that flips the z-component of the spin. Note that any 180° rotation around an axis in the xy-plane has this effect.

***J*-coupling.** The type of coupling present between two nuclear spins in a molecule in the liquid state.

Labeled molecule. A molecule in which some of the nuclei are substituted by less common isotopes. A common labeling for NMR QIP involves replacing the naturally abundant carbon isotope ^{12}C, with the spin-1/2 isotope ^{13}C.

Larmor frequency. The precession frequency of a nuclear spin in a magnetic field. It depends linearly on the spin's magnetic moment and the strength of the field.

Logical frame. The current frame with respect to which the state of a qubit carried by a spin is defined. There is an absolute (laboratory) frame associated with the spin observables σ_x, σ_y, and σ_z. The observables are spatially meaningful. For example, the magnetization induced along the x-axis is proportional to $\mathrm{tr}(\sigma_x|\psi\rangle\langle\psi|)$, where $|\psi\rangle$ is the physical state of the spin. Suppose that the logical frame is obtained from the physical frame by a rotation by an angle of θ around the z-axis. The observables for the qubit are then given by $\sigma_x^{(\mathrm{L})} = \cos(\theta)\sigma_x + \sin(\theta)\sigma_y$, $\sigma_y^{(\mathrm{L})} = \cos(\theta)\sigma_y - \sin(\theta)\sigma_z$, and $\sigma_z^{(\mathrm{L})} = \sigma_z$. As a result, the change to the logical frame transforms the physical state to a logical state according to $|\phi\rangle_{\mathrm{L}} = e^{i\sigma_z\theta/2}|\psi\rangle$. That is, the logical state is obtained from the physical state by a $-\theta$ rotation around the z-axis.

A resonant logical frame is used in NMR to compensate for the precession induced by the strong external field.

Magnetization. The magnetic field induced by an ensemble of magnetic spins.

The magnitude of the magnetization depends on the number of spins, the extent of alignment, and the magnetic moments.

Nuclear magnetic moment. The magnetic moment of a nucleus determines the strength of the interaction between its nuclear spin and a magnetic field. The precession frequency ω of a spin-1/2 nucleus is given by μB, where μ is the nuclear magnetic moment and B the magnetic field strength. For example, for a proton, $\mu = 42.7$ Mhz/T.

NMR spectrometer. The equipment used to apply rf pulses to and observe precessing magnetization from nuclear spins. Typical spectrometers consist of a strong, cylindrical magnet with a central bore in which there is a "probe" that contains coils and a sample holder. The probe is connected to electronic equipment for applying rf currents to the coils and for detecting weak oscillating currents induced by the nuclear magnetization.

Nuclear spin. The quantum spin degree of freedom of a nucleus. It is characterized by its total spin quantum number, which is a multiple of 1/2. Nuclear spins with spin 1/2 are two-state quantum systems and can therefore be used as qubits immediately.

Nutation. The motion of a spin in a strong z-axis field caused by a resonant pulse.

Nutation frequency. The angular rate at which a resonant pulse causes nutation of a precessing spin around an axis in the plane.

One-qubit quantum computing. The model of computation in which one can initialize any number of qubits in the state where Qubit 1 is in the state $|0\rangle_1$ and all the other qubits are in a random state. One can then apply one- and two-qubit unitary quantum gates and make one final measurement of the state of Qubit 1 after which the system is reinitialized. The model can be used to determine properties of the spectral density function of a Hamiltonian, which can be emulated by a quantum computer (Knill and Laflamme 1998).

Peak group. The spectrum of an isolated nuclear spin consists of one peak at its precession frequency. If the nuclear spin is coupled to others, this peak "splits," and multiple peaks are observed near the precession frequency. The nuclear spin's peak group consists of these peaks.

Precession. An isolated nuclear spin's state can be associated with a spatial direction with the help of the Bloch sphere representation. If the direction rotates around the z-axis at a constant rate, we say that it precesses around the z-axis. The motion corresponds to that of a classical top experiencing a torque perpendicular to both the z-axis and the spin axis. For a nuclear spin, the torque can be caused by a magnetic field along the z-axis.

Projective measurement. A measurement of a quantum system determined by a complete set of orthogonal projections whose effect is to apply one of the projections to the system ("wave function collapse") with a probability determined by the amplitude squared of the projected state. Which projection occurred is known after the measurement. The simplest example is that of measuring Qubit q in the logical basis. In this case, there are two projections, namely, $P_0 = |0\rangle_q {}^q\langle 0|$ and $P_1 = |1\rangle_q {}^q\langle 1|$. If the initial state of all the qubits is $|\psi\rangle$, then the probabilities of the two measurement outcomes 0 and 1 are $p_0 = \langle\psi|P_0|\psi\rangle$ and $p_1 = \langle\psi|P_1|\psi\rangle$, respectively. The state after the measurement is $P_0 = |\psi\rangle/\sqrt{p_0}$ for outcome 0 and $P_1 = |\psi\rangle/\sqrt{p_1}$ for outcome 1.

Pseudopure state. A state with deviation given by a pure state $|\psi\rangle\langle\psi|$.

Pulse. A transient field applied to a quantum system. In the case of NMR QIP, pulses are rotating magnetic fields (rf pulses) whose effects are designed to cause specific rotations of the qubit states carried by the nuclear spins.

Radio-frequency (rf) pulse. A pulse resonant at radio frequencies. Typical frequencies used in NMR are in this range.

Refocusing pulse. A pulse that causes a 180° rotation around an axis in the plane. A typical example of such a rotation is $e^{-i\sigma_x\pi/2} = -i\sigma_x$, which is a 180° *x*-rotation.

Resonant rf pulse. A pulse whose field oscillates at the same frequency as the precession frequency of a target nuclear spin. Ideally, the field is in the plane, rotating at the same frequency and in the same direction as the precession. However, as long as the pulse field is weak compared with the precession frequency (that is, by comparison, its nutation frequency is small), the nuclear spin is affected only by the corotating component of the field. As a result, other planar components can be neglected, and a field oscillating in a constant direction in the plane has the same effect as an ideal resonant field.

Rotating frame. A frame rotating at the same frequency as the precession frequency of a spin.

Rotation. In the context of spins and qubits, a rotation around σ_u by an angle θ is an operation of the form $e^{-i\sigma_u\theta/2}$. The operator σ_u may be any unit combination of Pauli matrices that defines an axis in three space. In the Bloch sphere representation, the operation has the effect suggested by the word "rotation."

Spectrum. In the context of NMR, the Fourier transform of an FID.

Weak measurement. A measurement involving only a weak interaction with the measured quantum system. Typically, the measurement is ineffective unless an ensemble of these quantum systems is available so that the effects of the interaction add up to a signal detectable above the noise. The measurement of nuclear magnetization used in NMR is weak in this sense.

Realizing a Noiseless Subsystem in an NMR Quantum Information Processor

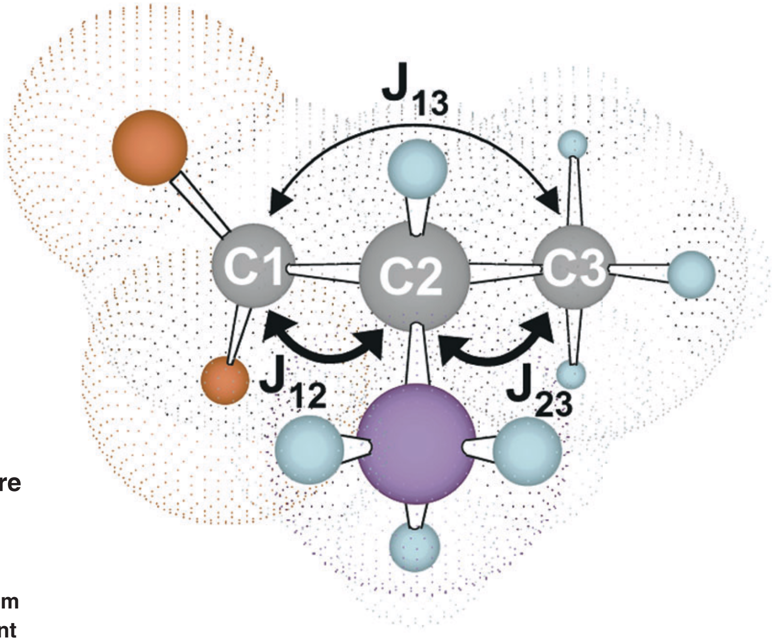


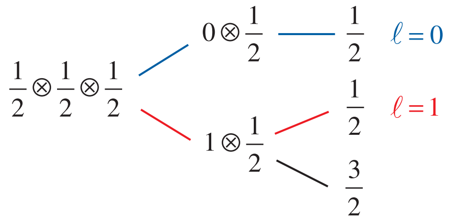
Figure 1. Molecular Structure of ^{13}C-Labeled Alanine
The diagram shows the three carbon-13 (^{13}C) spins used as qubits in the noiseless subsystem experiment as well as the relevant *J*-couplings between those qubits.

Lorenza Viola and Evan M. Fortunato

The fact that the occurrence of symmetries in a physical system generally implies the existence of conserved quantities and that these symmetries can be exploited to ease the understanding of the system's behavior is a well-known lesson in physics. The notion of a noiseless subsystem (NS) (Knill et al. 2000) captures this lesson in the context of quantum information processing (QIP), where the challenge is to protect information against the detrimental effects of noise. The link between symmetries, conserved quantities, and NS was discussed at length on page 216 of the article "Introduction to Error Correction". The essential message is that, by encoding information into an abstract subsystem that corresponds to a preserved degree of freedom, noiselessness is guaranteed even if errors still evolve the overall system's state.

Here, we focus on the NS of three spin-1/2 particles introduced in the above-mentioned article (see page 201), along with a discussion of the error-correcting properties of this NS. The physical system is composed of three qubits, subjected to a "far-field" interaction with the environment, whereby the latter couples to the qubits without distinguishing among them. The resulting collective-noise model involves all possible error operators that are symmetric under permutation of the three particles and is specified in terms of the error generators $J_u = (\sigma_u^{(1)} + \sigma_u^{(2)} + \sigma_u^{(3)})/2$, where $u = x, y, z$. By recalling the meaning of the single-spin Pauli operators σ_u^k, the observable J_u represents the projection of the total spin angular momentum $\boldsymbol{J}$ along the u-axis. Because the total-spin observable $J^2 = \boldsymbol{J} \cdot \boldsymbol{J}$ commutes with the error generators and z defines the quantization axis, the eigenvalues j and j_z of J^2 and J_z, respectively, provide useful quantum numbers to label basis states for the three particles.

The NS of interest resides in the four-dimensional subspace $\mathcal{H}_{1/2}$ of the states carrying total angular momentum $j = 1/2$ and having a total z-component $j_z = \pm 1/2$. However, specifying j and j_z does not suffice for completely labeling the states in $\mathcal{H}_{1/2}$: An additional quantum number is needed for removing the two-dimensional degeneracy that remains. Physically, this degeneracy simply means that there are two distinct paths for obtaining a total angular momentum $j = 1/2$ out of three elementary 1/2 angular momenta:



Let the additional quantum number $\ell = 0, 1$ label the two possible routes in the above diagram. Because collective noise does not distinguish among the individual spins and the final eigenvalue j is the same for both paths, the noise can neither distinguish the realized value of ℓ nor change that value. This conserved quantum number can be directly related to the eigenvalues $s_z = \pm 1$ of the $\sigma_z^{(NS)}$ observable of a noiseless qubit, $s_z = 2\ell - 1$. In general, noiseless qubit operators will remain invariant under rotations. The simplest scalars under the rotations are the dot products $s_{12} = \sigma^{(1)} \cdot \sigma^{(2)}$, $s_{23} = \sigma^{(2)} \cdot \sigma^{(3)}$, and $s_{31} = \sigma^{(3)} \cdot \sigma^{(1)}$.

Thus, $\sigma_u^{(NS)}$ observables for the noiseless qubit, where $u = x, y, z$, can be constructed by combining s_{12}, s_{23}, s_{31}, and the identity into three operators that "behave like" the Pauli matrices (Viola et al. 2001a). A good choice is given by $\sigma_x^{(NS)} = 1/2(\mathbb{1} + s_{23})$, $\sigma_y^{(NS)} = \sqrt{3}/6(s_{31} - s_{12})$, and $\sigma_z^{(NS)} = i\sigma_y^{(NS)}\sigma_x^{(NS)}$, where projection onto the relevant $\mathcal{H}_{1/2}$ subspace is understood. Note that the action corresponding to $\sigma_x^{(NS)}$ is simply a permutation exchanging the last two spins. (For an alternative construction of the NS observables, see the article "Introduction to Error Correction," page 216.) Identifying the NS through its observables is equivalent to identifying it through the explicit state space correspondence given in Equation (28) of the above-mentioned article.

The experimental implementation of the three-qubit NS (Viola et al. 2001b) was performed with liquid-state NMR techniques. The three spin-1/2 carbon nuclei of carbon-13-labeled alanine were used as qubits (Figure 1). The information to protect is an arbitrary one-qubit state, $|\psi\rangle = a|0\rangle + b|1\rangle$, where a and b are arbitrary complex amplitudes, and $\langle\psi|\psi\rangle = 1$. This information is initially stored in spin 3, meaning that the three carbon spins are initialized in a pseudopure state corresponding to $|0\rangle_1|0\rangle_2|\psi\rangle_3 = |00\psi\rangle = a|000\rangle + b|001\rangle$. A unitary transformation U_{enc} encodes this input state into a superposition of the two basis states in $\mathcal{H}_{1/2}$ with $j = 1/2$ and $j_z = -1/2$. That is,

$$U_{\text{enc}}|00\psi\rangle \leftrightarrow a|\downarrow\rangle \cdot |0\rangle + b|\downarrow\rangle \cdot |1\rangle = |\downarrow\rangle \cdot |\psi\rangle \quad , \tag{1}$$

where the subsystem representation of Equation (28) has been used.

The three qubits remain stored in the NS memory for a fixed evolution period t_{ev}, during which errors can occur. In a given set of experiments, these errors are designed to implement a desired collective-noise process $\mathcal{E}_{\text{coll}}$ described by a set of error operators $\{E_a\}$. Because of their collective nature, these errors affect only the syndrome subsystem in the pair. Finally, following the evolution period, the unitary transformation U_{dec} decodes a generic noisy state $E_a(|\downarrow\rangle \cdot |\psi\rangle)$ in $\mathcal{H}_{1/2}$ back to the computational basis.

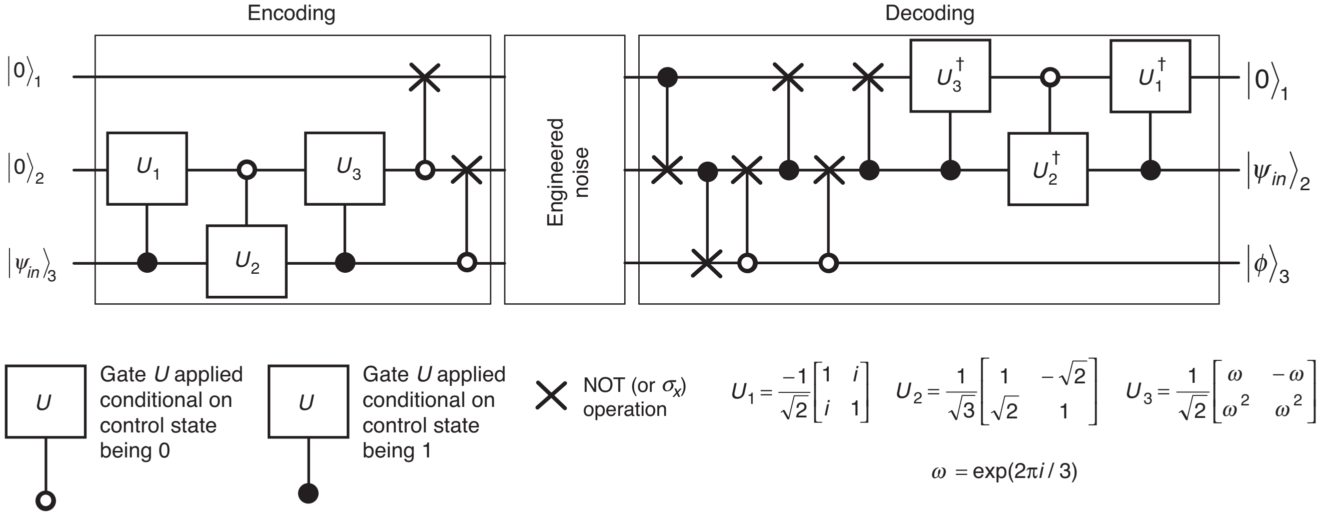

Encoding
Decoding
$|0\rangle_1$
$|0\rangle_2$
$|\psi_{in}\rangle_3$
U_1
U_2
U_3
Engineered noise
$U_3^\dagger$
$U_2^\dagger$
$U_1^\dagger$
$|0\rangle_1$
$|\psi_{in}\rangle_2$
$|\phi\rangle_3$
U
Gate U applied conditional on control state being 0
U
Gate U applied conditional on control state being 1
NOT (or σ_x) operation
$U_1 = \frac{-1}{\sqrt{2}}\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$
$U_2 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & -\sqrt{2} \\ \sqrt{2} & 1 \end{bmatrix}$
$U_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} \omega & -\omega \\ \omega^2 & \omega^2 \end{bmatrix}$
$\omega = \exp(2\pi i/3)$

Figure 2. Logical Quantum Network

The diagram shows a logical quantum network for the three-qubit NS experiment. The logical manipulations were translated into sequences of radio-frequency pulses and delays, and complete pulse programs for U_{enc} and U_{dec} resulted from the compilation of the partial pulse programs for individual gates. The pulses were designed to ensure self-refocusing of all the unwanted *J*-coupling and chemical-shift evolutions.

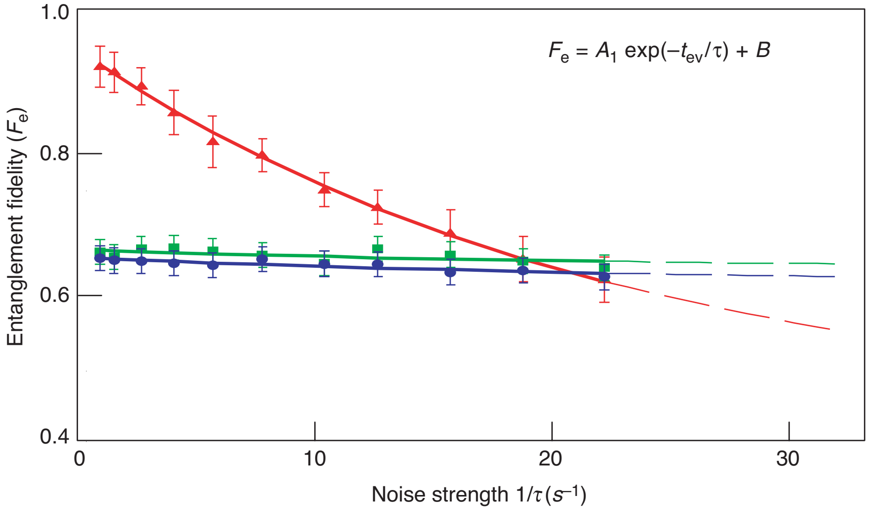
This procedure has the effect of returning the quantum state $|\psi\rangle$ onto qubit 2 upon discarding ("tracing over") spins 1 and 3,

$$\mathrm{Tr}_{1,3}\{U_{\mathrm{dec}}[\mathcal{E}_{\mathrm{coll}}(|\downarrow\rangle\langle\downarrow| \cdot |\psi\rangle\langle\psi|)]U^{-1}_{\mathrm{dec}}\} = |\psi\rangle_2\langle\psi| \ . \tag{2}$$

Figure 2 is a sketch of the quantum network for the experiment.

During the delay period between encoding and decoding, we use gradient diffusion techniques to engineer a desired collective-noise process. In order to fully explore the robustness properties of information encoded in the NS, we applied various error models corresponding to noise along a single axis (see Figure 3), as well as more complicated double- and triple-axis noise processes obtained by "cascading" the action of error models along different spatial directions, in sequence, within a single evolution period (see Table I). To quantify the accuracy of the implemented NS in preserving the quantum data $|\psi\rangle$, we experimentally extracted the entanglement fidelity F_e of the overall process (including encoding, decoding, and engineered noise during storage), where $F_e = 1$ implies perfect preservation.

Our results in Figure 3 and Table I indicate that, as expected, the effects of the applied noise increase exponentially as a function of noise strength for unencoded (UN) information but are largely independent of noise strength for information encoded in the NS. That independence demonstrates that the NS functions as an "infinite-distance" quantum error-correcting code for arbitrary collective errors. On the other hand, the F_e is always about the same and less than 1 in all the NS experiments. The constant reduction in fidelity is suggestive of errors introduced during encoding and decoding manipulations, as well as of noise due to natural noncollective relaxation processes during the whole experiment. ■



$F_e = A_1 \exp(-t_{ev}/t) + B$

A_1	B	Quantum Process
0.51 ± 0.04	0.43 ± 0.03	▲ UN, *y*-axis noise
0.03 ± 0.03	0.64 ± 0.02	■ NS-encoded, *y*-axis noise
0.03 ± 0.03	0.62 ± 0.02	● NS-encoded, *z*-axis noise

Figure 3. Entanglement Fidelities for Engineered Collective Noise along a Single Axis

The fidelity of UN information subjected to engineered collective noise along the *y*-axis (red) decreases exponentially with noise strength τ^{-1} whereas the fidelity of NS-encoded information subjected to collective noise along either the *y*-axis (green) or the *z*-axis (black) remains almost constant independent of noise strength.

In each case, noise was applied for a fixed evolution period t_{ev} of approximately 44 ms.

The flatness of the curve interpolating the NS data demonstrates the behavior of the NS as an infinite-distance quantum error-correcting code for single-axis collective errors of arbitrary strength. The smooth fits to the data are derived from the exponential and parameters displayed under the figure.

Table I. Entanglement Fidelities for Engineered Collective Noise along Two and Three Axes

Quantum Process	Entanglement Fidelity (F_e)
Q_{zx}^{UN}	0.24
Q_{00}^{NS}	0.70
Q_{zx}^{NS}	0.70
Q_{zy}^{NS}	0.70
Q_{000}^{NS}	0.67
Q_{yzx}^{NS}	0.66

Q stands for the one-qubit processes implemented during each run.

Superscripts tell whether the system has been encoded or not.

Subscripts *zx*, *zy*, and *yzx* are for the axes along which noise processes with maximum achievable strength were applied in cascade. Subscripts 00 and 000 indicate that no noise was applied.

Two subscripts indicate shorter delay periods than three subscripts.

Statistical uncertainties in all F_e values are approximately 2%.

Further Reading

Knill, E., R. Laflamme, and L. Viola. 2000. Theory of Quantum Error Correction for General Noise. *Phys. Rev. Lett.* **84**: 2525.

Viola, L., E. Knill, and R. Laflamme. 2001a. Constructing Qubits in Physical Systems. *J. Phys. A* **34** (35): 7067.

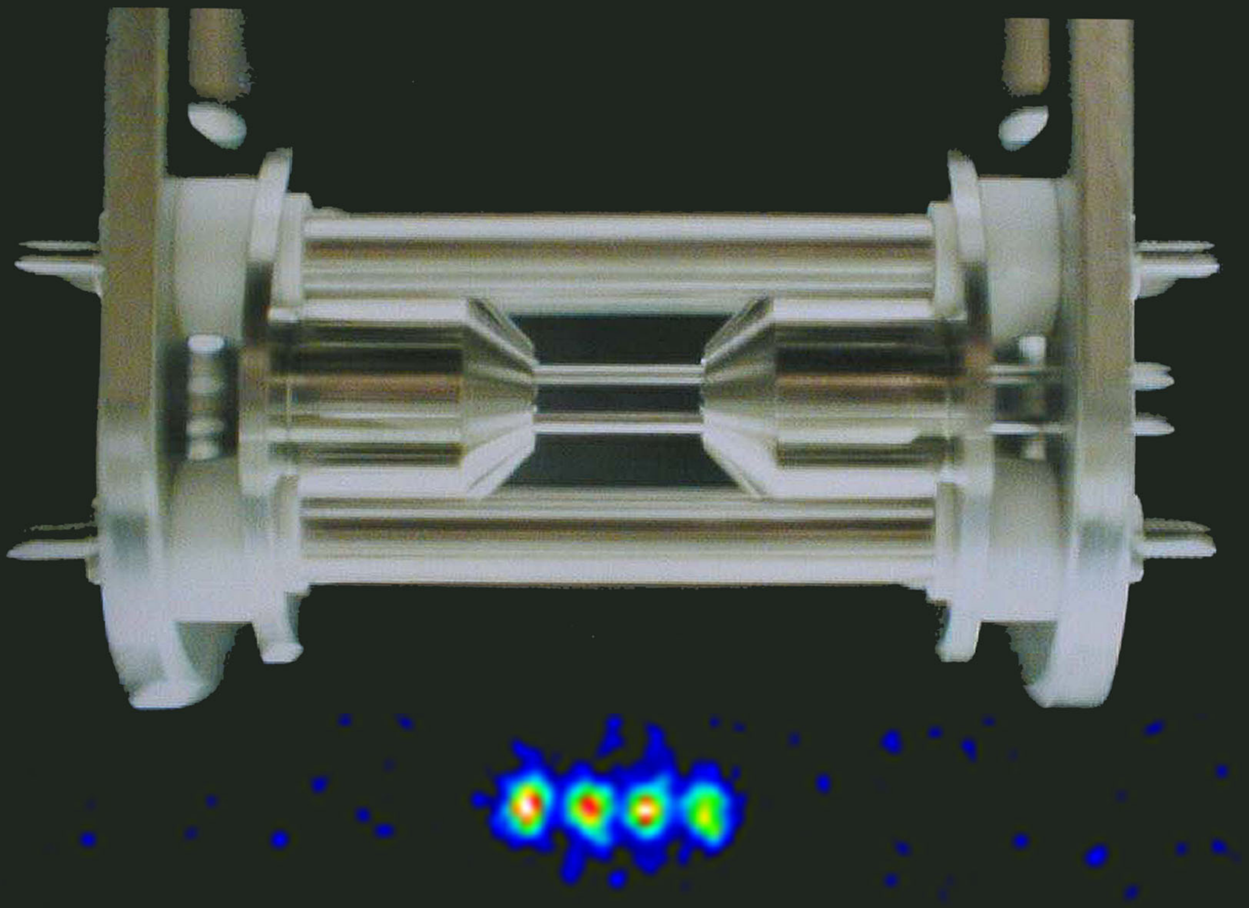
Viola L., E. M. Fortunato, M. A. Pravia, E. Knill, R. Laflamme, and D. G. Cory. 2001b. Experimental Realization of Noiseless Subsystems for Quantum Information Processing. *Science* **293**: 2059.

Lorenza Viola obtained a Ph.D. degree in physics from the University of Padova (Italy) in 1996. After being a postdoctoral fellow at the D'Arbeloff Laboratory at the Massachusetts Institute of Technology, Lorenza became a Director-funded and then a J. R. Oppenheimer postdoctoral fellow at Los Alamos. Her recent research has been focused on quantum information science, with emphasis on devising schemes for controlling noisy quantum devices and performing robust quantum computation. Lorenza has been a key contributor to the development of quantum error-suppression techniques based on dynamical decoupling and to the theoretical characterization and experimental verification of the notion of a noiseless subsystem as the most general approach to noise-free information storage.

> "…it seems that the laws of physics present no barrier
> to reducing the size of computers until bits are the size of atoms,
> and quantum behavior holds dominant sway."
> —R. P. Feynman, 1985

Ion-Trap Quantum Computation

Michael H. Holzscheiter



Quantum computation requires a very special physical environment. Numerous operations must be performed on the quantum states of the qubits (or quantum bits) before those states decohere, or lose the interlocking phase relationships that give quantum computation its power. Thought to be an unavoidable outcome of the interaction between the quantum state and the environment, decoherence threatens the life of a quantum system.

Any attempt at building a real quantum computer therefore leads to what some scientists refer to as the yin-yang of quantum computation: On the one hand, the qubits must interact weakly with the environment in order to limit decoherence. On the other hand, they must be easily accessible from the outside and must interact strongly with each other, or else we could not manipulate the quantum state, implement quantum algorithms, and read out the result of a calculation in a timely fashion. How can we hope to meet such contradictory goals?

Ion-trap quantum computers, as originally proposed by Ignacio Cirac and Peter Zoller (1995), offer a possible solution to this dilemma. As its name implies, an ion trap confines charged particles to a definite region of space with magnetic and electric fields. In a specific realization of such a trap, called a linear radio-frequency quadrupole (RFQ) trap, or a linear Paul trap (Raizen et al. 1992, Walther 1994), time-varying electric fields are used to hold a line of ions in place—like pearls on a string. These ions serve as the physical qubits of the quantum computer. Immobilized by the trapping fields and confined inside an ultrahigh vacuum chamber, they are effectively isolated from the environment. However, by addressing individual ions with sharply defined laser beams, we can initialize the computer, control the qubit states during the operation of logic gates, and read out the results at the end of the computation. The interaction between the individual ions is mediated by the Coulomb force between the charged particles.

This article discusses the design principles for isolating single ions in a linear Paul trap (Paul et al. 1958), whose operational principles are described in detail. The individual elements of an ion-trap computer will be introduced, and how to initialize, manipulate, and interrogate the qubits will be explained. Specific schemes that were implemented in the quantum computation project at Los Alamos (Hughes et al. 1998) will illustrate the descriptions. Ion-trap quantum computation is rapidly evolving, and numerous groups around the world are developing new ideas and experimental techniques. The reader will get a flavor of this activity in the last section of the article, which summarizes several important results achieved within the last few years: the on-demand creation of entangled states of up to four ions by the National Institute of Science and Technology (NIST) group in Boulder, Colorado; the development of a novel cooling scheme by a group at the University of Innsbruck, Austria, which would allow researchers to quickly cool large numbers of trapped ions with drastically reduced operational overhead; and the construction of an effective defense against the forces of decoherence.

The Physics of Ion Traps

Two basic types of devices can confine charged particles to well-defined regions of free space: Penning traps and Paul traps. The Penning trap, which was primarily developed by Hans Dehmelt at the University of Washington in Seattle, uses a strong magnetic field and a static electric field to create a nearly perfect three-dimensional, harmonic trapping potential (Dehmelt 1967). Some of the most precise tests of fundamental physical symmetries to date have been conducted with this device, whose operating principles are described in the box "The Penning Trap" on the next two pages.

Although Penning traps nicely solve the fundamental problems of ion confinement, so far they have not been used for quantum computation. The trap's strong magnetic field causes ions to move rapidly in a circle (the cyclotron motion discussed in the box), whereas we want the physical qubits to have as little motion as possible. That is why the favored trap for quantum computation is the Paul trap, in which there is no magnetic field and oscillating electric fields (as opposed to static ones) confine the ions. This device was invented by Wolfgang Paul from the University of Bonn in Germany, who shared the 1989 Nobel Prize in physics with Dehmelt.

Paul enjoyed telling the following anecdote about how he hit upon the idea for his device. Germans like soft-boiled eggs for breakfast, and on a particular Sunday morning, Paul had prepared two eggs of different sizes and had placed them on a serving tray. When he started to walk, tray in hand, toward the bedroom to surprise his wife with breakfast in bed, the eggs began to roll. He counteracted their motion by shaking the tray and was able to confine the larger egg to the center by shaking with a particular frequency and amplitude. (It was certainly not a well-defined harmonic shaking.) The smaller egg, however, kept rolling toward the edge, so Paul changed amplitude and frequency and successfully prevented this egg from falling, at the expense of allowing the larger one to wobble toward the edge. Whether he ever reached the bedroom with both eggs on the tray and enjoyed a leisurely breakfast with his wife remains unknown, but that morning Paul realized not only the basic principle of the RFQ trap but also the mass-selective feature of such an instrument. At that time, he was keen on developing a mass filter for ions, that is, a two-dimensional structure that could transmit an ion with a specific charge-to-mass ratio and not any other ratio. Eventually, Paul's idea was used to generate three-dimensional, mass-selective confinement systems, but Cirac and Zoller returned to the original two-dimensional structure and proposed using it as the basis for a quantum computer.

The Linear Paul Trap. To understand the linear RFQ trap, consider a positively charged ion floating in free space and surrounded by four infinitely long conducting rods, as shown in Figure 1. We can give one pair of opposing rods a positive charge and the other pair a negative charge

Continued on page 268

The Penning Trap

Decades before individual ions were considered as candidates for qubits in a quantum computer, experimental physicists were challenged to realize a simpler Gedanken, or thought, experiment embodied by the statement often made by theorists: "Consider a single (silver) ion in a uniform magnetic field" (Tanoudji et al. 1977). Thought became reality in 1973, when Hans Dehmelt and his colleagues at the University of Washington in Seattle were able to capture a single charged particle in a Penning trap. The ion that drifted into the central region of that device was trapped by a strong, uniform magnetic field and by the electrostatic field produced by a set of specially shaped electrodes. The entire device operated under ultrahigh vacuum to limit the interactions between the ion and the background atoms.

The University of Washington group refined the technique and used the trap to confine a single electron (Wineland et al. 1973) and later a single barium ion, using an RFQ Paul trap (Neuhauser et al. 1980), and performed precision spectroscopy on these systems. The special arrangement of fields caused the single electron to behave as if it were bound to a nucleus for it displayed a set of energy levels, or excited states, similar to those of the hydrogen atom. Dehmelt therefore named his electron in a Penning trap "geonium—a single electron bound to Earth." The artificial geonium "atom" was, in a sense, closer to perfection than a real atom. The spacing between energy levels was nearly constant because it reflected the trap's nearly perfect harmonic-oscillator potential. Dehmelt and coworkers used geonium to perform some of the most precise tests of fundamental symmetries. In a more mundane fashion, Dehmelt called the ion ASTRID (for "a single trapped ion dancing"). (Perhaps, if you keep an ion or electron for such a long time, you may become attached to it.)

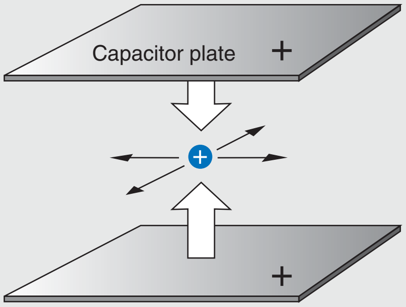


Figure A. Electrostatic Forces
The positively charged particle is repulsed by the capacitor plates but is free to move anywhere in the horizontal plane.

To understand the operating principles of the Penning trap, consider a charged particle (ion) moving freely in space. To confine it to a specific spot in space, we can apply electrical forces to its charge. If we place the ion between two parallel conducting plates that are charged to an electric potential of the same sign as the ion, the Coulomb repulsion will keep the particle from moving closer to either plate (see Figure A).

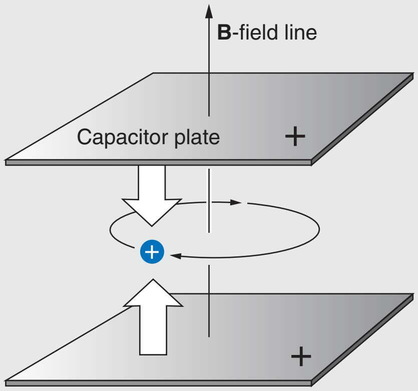


Figure B. Applying a Magnetic Field
A magnetic field causes the ion to circle around a field line (cyclotron motion), thus confining the ion in the horizontal plane.

The ion can still move in directions parallel to the conductors. We can try to remove all escape routes by placing more conductors around the particle. But Michael Faraday discovered more than 150 years ago that an electric field cannot penetrate a closed metal enclosure—hence, the penchant for science museums to place a person inside a "Faraday cage" that is then exposed to violent lightning bolts. The courageous volunteer remains unharmed because the lightning's electric field vanishes inside the cage. Similarly, if we fully enclose our particle in a cage of conducting plates, the electric field disappears, and we lose the forces holding the particle from the walls.

A more successful approach is to use the fact that an electric charge moving in a magnetic field will experience a force perpendicular to the direction of both the magnetic field and the particle's velocity (the Lorenz force $\boldsymbol{F} = q\mathbf{v} \times \mathbf{B}$). Therefore, if we apply a magnetic field perpendicular to our parallel conducting plates (see Figure B), we force the ion onto a circular path around the magnetic field line, closing off the sideway escape routes.

Whereas the system shown in Figure B can confine charged particles (and has been used for a number of experiments), the special character of the Penning trap is given by the clever shaping and arrangement of the electrodes. As shown in Figure C, two end caps shaped as hyperbolae of revolution replace the parallel plates, and a ring-shaped center electrode defines the electrostatic potential on the edge of the trap.

The arrangement shown in Figure C not only leads to perfect confinement of individual charged particles but also allows the motion of a trapped particle to be separated into three independent harmonic motions. In order of decreasing frequency, the three motions are (1) the fast "cyclotron" motion of the charge around the magnetic field lines, (2) a slower oscillation in the direction of the magnetic field that is due to the electrostatic repulsion from the two end caps, and (3) a much slower drift motion that is due to the crossed electric and magnetic fields (see Figure D).

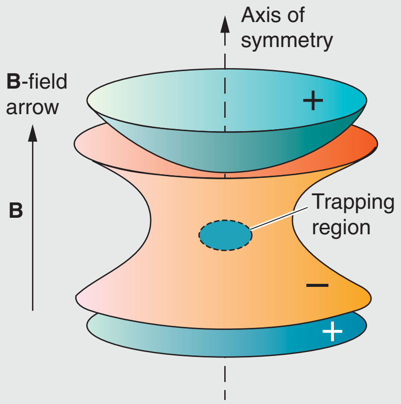


Figure C. The Penning Trap
The two endcaps, which are hyperbolae of revolution, replace the flat capacitor plates. The central ring electrode helps define the harmonic potential at the center of the trap.

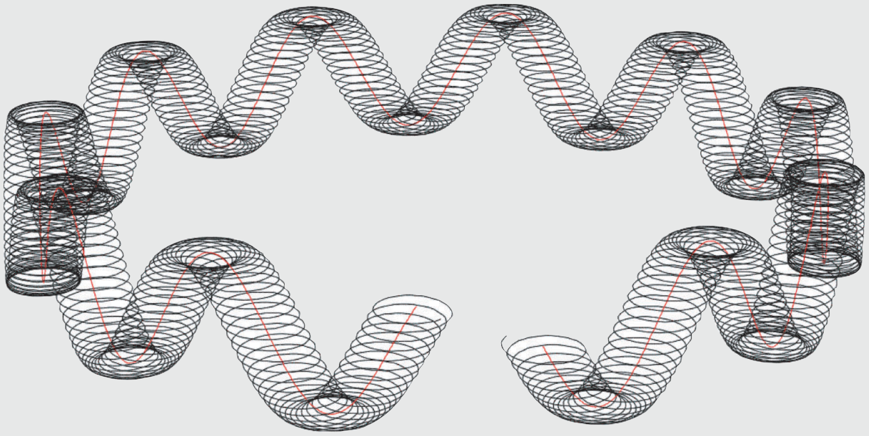


Figure D. Motion in the Penning Trap
The three-dimensional motion of an ion in the trap consists of three harmonic motions: a fast cyclotron motion, a slower up-down oscillation, and a slow circular drift motion.

The drift motion is easily understood if one focuses on the cyclotron motion. The positively charged particle is accelerated toward the negatively charged ring electrode as it moves away from the electrical center of the trap. This acceleration increases the radius of curvature for the outer half of the cyclotron motion. As the particle moves back toward the center during the second half of the cyclotron motion, it decelerates, and the radius of curvature decreases. The net effect is a distortion of the circular cyclotron motion into a spiral that bends around the electrical center of the trap, as seen in Figure E.

The harmonic motions account for the almost constant spacing between energy levels in Dehmelt's geonium atom, but this orderliness is hardly noticeable in the roller-coaster-like motion of a trapped particle. If you actually want to experience the particle's motion yourself, there is a carnival ride in which these three components of motion are present—but watch your stomach!

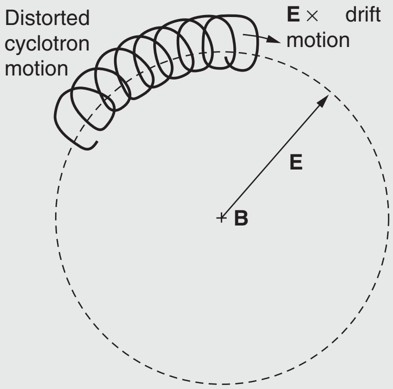


Figure E. Drift
Schematic of the drift motion that results from the crossed electric (E) and magnetic (B) fields.

Continued from page 265

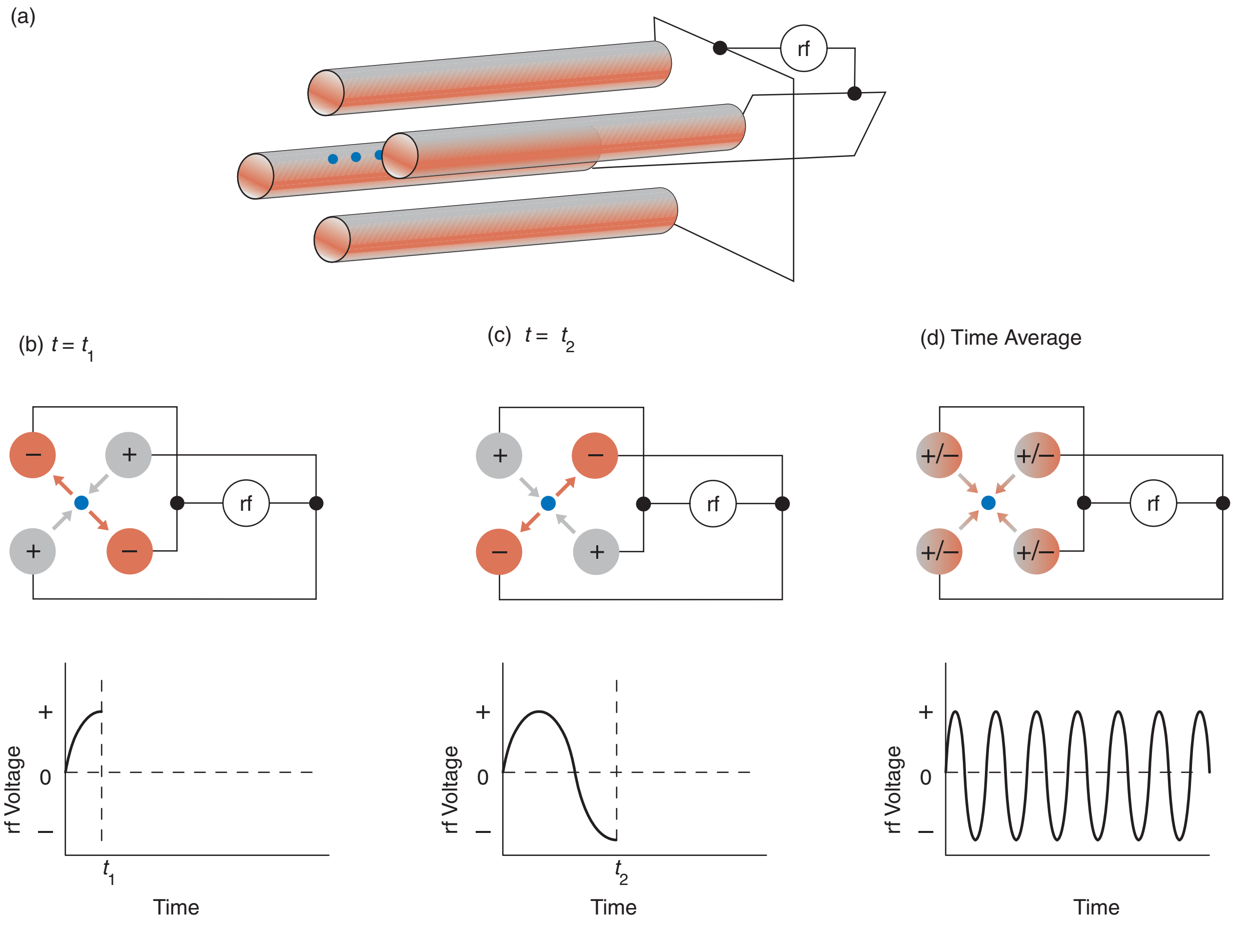


Figure 1. Principles of the Linear Paul Trap

(a) The linear Paul trap consists of four conducting rods. Two opposing rods are connected to one pole of a radio-frequency (rf) voltage source, whereas the remaining two are connected to the other pole. The axis of symmetry between the rods is the trap axis. (b) With the rods charged as shown, the resulting electric force pushes a positive ion to the negative rods and repels it from the positive ones. (c) Half an rf period later (see graph below), the polarity of all rods is reversed, and the direction of the force also reverses. (d) If the polarity changes fast enough, a heavy ion becomes stuck in a rapid back-and-forth motion. Because the electric fields are at a minimum at the trap axis, an effective force pushes the ion toward the center, where it becomes trapped (although it is still free to move along the axis). The blue dots seen between the rods in part (a) represent a string of radially trapped ions. The string can be confined axially when a positively charged electrode (end cap) is placed at each end of the rods.

(relative to some arbitrary “zero” potential).[1] The positive ion feels a repulsive force from the positively charged conductors and is pushed toward the center of the trap. The ion simultaneously feels an attractive force due to the negatively charged conductors and is pulled outwards.

If we now reverse the polarity of our four electrodes, interchanging plus for minus and minus for plus, the ion’s motion will begin to reverse. Where it was moving out, it will now be moving in, and vice versa. However, if the reversal takes place quickly, the “heavy” ion cannot easily respond; it has too much inertia to follow the fast changes in the electric field exactly. Instead, the ion will respond to the time-averaged electric field. If we switch the polarity of the electrodes at a few megahertz (or a few million times a second) by applying a radio-frequency (rf) voltage and if the amplitude is correct, then the time-averaged field generates a harmonic pseudopotential with its minimum located at the trap axis. The ion is pushed to the bottom of the pseudopotential well and becomes trapped forever—at least in principle—near the center.

---

[1] For reasons discussed on the previous two pages, all four conductors cannot be positively charged, or the electric fields within the trap would disappear.

To generate the mass selectivity sought by Paul, we add a positive direct-current (dc) component to the rf voltage. Positive ions outside a certain mass range feel less of a restoring force from the pseudopotential and are kicked out of the trap by the repulsive dc field. This technique is one of several that we can use to preferentially retain qubit ions instead of, say, a residual gas ion that may be present in the ultrahigh vacuum trap.

Of course, the ion's motion is still unrestricted in the direction parallel to the trap's axis. For confinement in this third dimension, we simply add an electric dc potential to a pair of "end electrodes" that are on either side of the region of interest. This axial field plugs up the escape route along the symmetry axis of the system, and the ion becomes trapped in three dimensions. By substantially reducing the ion's energy (cooling), we coax the ion into lying along the central portion of the trap axis, where the radial and axial confining potentials are at a minimum. If several very cold ions are in the trap, then they all fall to the center, and the mutual Coulomb repulsion between the ions causes them to line up neatly along the axis.

Motion in the Trap. Although the combination of rf and dc fields within the trap drives the ion into a complex radial motion, that motion is fully described by a set of differential equations called "Mathieu's Equations." The bound solutions of those equations yield a stability diagram that allows one to evaluate the effectiveness of the trap, given the values of several critical parameters, such as the amplitudes of the rf and dc components of the voltage, the rf, the ion mass, and the size of the trap (Dawson 1976).

As long as we keep the values of the critical operational parameters within certain ranges, an ion will remain bound to the axis of the device. Furthermore, the magnitude of the restoring force of the pseudopotential holding the ions in the radial direction will remain directly proportional to the distance from the center—the hallmark of a harmonic potential.[2] In other words, to a good approximation, the ions will undergo harmonic oscillations in the radial direction with frequency ω_r (or with frequency ω_x and ω_y in case the potential is different in the *x*- and *y*-directions). This motion is commonly referred to as the secular motion.

The ions' motion can become distorted if the minima of the rf field and the pseudopotential are misaligned within the trap. Misalignment can occur because of some small asymmetry in the trap's construction or because of small dc patch potentials on the electrode surfaces. Regardless of the reason, misalignment will cause the ions to lie "off center" (that is, off the line where the rf field vanishes). Those ions will experience the strong gradient of the rf field and undergo rapid oscillations—at the frequency of the applied rf field—around their time-averaged equilibrium position. This so-called micromotion is the main source of ion heating. We can suppress the micromotion by adding a compensating dc voltage to some of the rf electrodes (or to auxiliary control electrodes) and thereby shift the ions' positions closer to the actual rf center.

In addition to exhibiting secular motion and the unwanted micromotion, an ion or, more important, a string of ions will also vibrate in the axial direction. The motion will be harmonic because the dc voltage on the end electrodes creates a harmonic potential along some length of the trap axis. The vibrations are similar to those exhibited by a set of pendula connected to each other by springs; the swinging of one pendulum sets the others in motion (see Figure 2). Unlike vibrations of classical pendula, however, the vibrations exhibited by a string of ions are quantized; that is, the amplitude of the motion depends on the number of quanta (phonons) in the vibrational mode.

For *N* ions in a trap, there are *N* axial vibrational modes and an additional 2*N* modes for motions transverse to the axis. Each mode has a distinct vibrational frequency. The lowest-frequency (lowest-energy) vibration is the so-called common mode, in which the ions oscillate back and forth in unison along the axis. This mode figures heavily in the original quantum-computing scheme of Cirac and Zoller. Because all ions participate in the common-mode oscillation, when we add (or remove) a quantum of energy to this motion by interacting with one of the ions, we influence all other ions in the string. Any two qubits, regardless of the distance between them, can therefore be coupled together to perform logic operations.

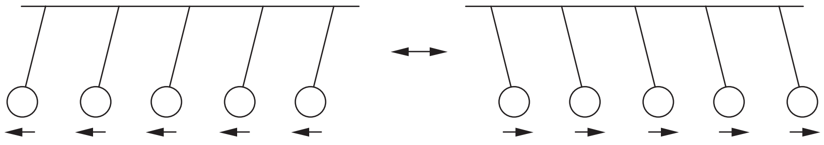
Other proposals make use of some of the higher-frequency modes to couple qubits together (James 1998a). These modes have more-complex vibrational patterns and relatively higher excitation energies than the common mode, but it still takes very little energy to excite them. Even a string of very cold ions will vibrate in some intricate expression of the various modes, a problem that is addressed in the discussion of ion cooling.

If only a few ions are confined in the trap, the ions will align themselves linearly along the axis. But increasing the number of ions or increasing the dc voltage applied to the end electrodes introduces instabilities because

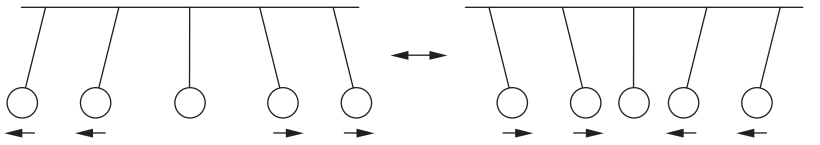
---

[2] To generate a pure harmonic potential, the four electrodes should have hyperbolic cross sections, but in practice we approximate that ideal shape with cylindrical rods.

(a) Common Mode



(b) Breathing Mode



(c) Higher-Order Arial Mode

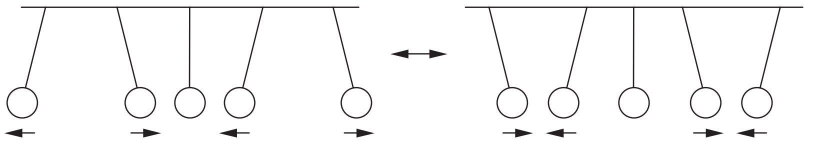


Figure 2. Vibrational Modes

A set of strongly coupled pendula can be used to envision the vibrational motion of a string of ions in a harmonic potential. These vibrational modes affect all ions simultaneously. If we set any one of the pendula in motion, the others will move. Similarly, if we grab hold of any pendulum and stop it, all others will stop. (a) The common mode (center-of-mass mode), in which all pendula swing one way and then the other, has the lowest frequency (lowest energy) of all modes. Using this mode to couple two qubits together in the trap is the basis of the Cirac-Zoller proposal. (b) The breathing mode, in which pendula at opposite ends move in opposite directions, has the next highest frequency. For an odd number of pendula, the middle one does not move. This mode is less susceptible to heating by external noise sources and has also been used to couple qubits. (c) Shown here is another higher-order mode. In an ion trap, the ions can vibrate in three dimensions; for *N* trapped ions, there are 3*N* vibrational modes.

the ions are effectively squeezed closer together. The Coulomb repulsion between neighboring ions becomes stronger than the radial restoring force, and the ions start buckling out into a zigzag pattern. When even more ions are added, the zigzag pattern develops into a complex three-dimensional helical structure (Raizen et al. 1992, Walther 1991, 1994). Some of the ions will move away from the axis and will experience the strongest micromotion heating—a situation clearly to be avoided. We have studied this transition from linear to three-dimensional structures in some detail (Enzer et al. 2000) and have quantified the parameter space available for quantum computing in a linear RFQ ion trap.

Elements of the Ion-Trap Quantum Computer

In 1994, inspired by the great success of ion traps in the field of precision measurements, Cirac and Zoller proposed that the RFQ ion trap had the right characteristics to support the long sequence of precision operations required for quantum computation. A string of ions trapped along the symmetry axis of the trap would be the quantum register of the computer. Each ion could be addressed by tightly focused laser beams, initialized to an arbitrary state, manipulated, and then probed at the end of a calculation. Most important, the isolation from the environment afforded by the trap would allow for long coherence times.

One- and Two-Qubit Operations. The logical qubit states $|0\rangle$ and $|1\rangle$ of the ion-trap quantum computer must be defined as they always are for any quantum computer. (To stress that the 0 and 1 used to designate the states are notational and have no numerical meaning, we have used a font different from the one for the numbers 0 and 1.) We simply identify the ion's electronic ground state with the qubit state $|0\rangle$ and a long-lived excited state with the qubit state $|1\rangle$.

We also want to apply a unitary transformation to a single qubit, that is, to implement a one-qubit gate, and rotate the qubit in Hilbert space to an arbitrary superposition of the $|0\rangle$ and $|1\rangle$ states. (Two-level systems and the rotation of a qubit in Hilbert space are discussed in the article "Quantum Information Processing" on page 2.) To do so, we subject the ion to a laser pulse of a specific amplitude, frequency, and duration. Assuming the ion is in its ground state, the laser pulse will cause the electron wave function of the target ion to evolve to some superposition of the ground and excited states. (We cause the electron to undergo part of a Rabi oscillation.) Illuminating the ion with a so-called π-pulse, for example, will evolve the electron wave function through half a Rabi oscillation period and leave the ion in the excited state. The qubit would have rotated from the $|0\rangle$ to the $|1\rangle$ state. If the duration of the pulse is halved (a so-called $\pi/2$-pulse), the ion would be left in an equally weighted superposition of the ground and excited states. The qubit would have rotated to the $1/\sqrt{2}(|0\rangle + |1\rangle)$ state.

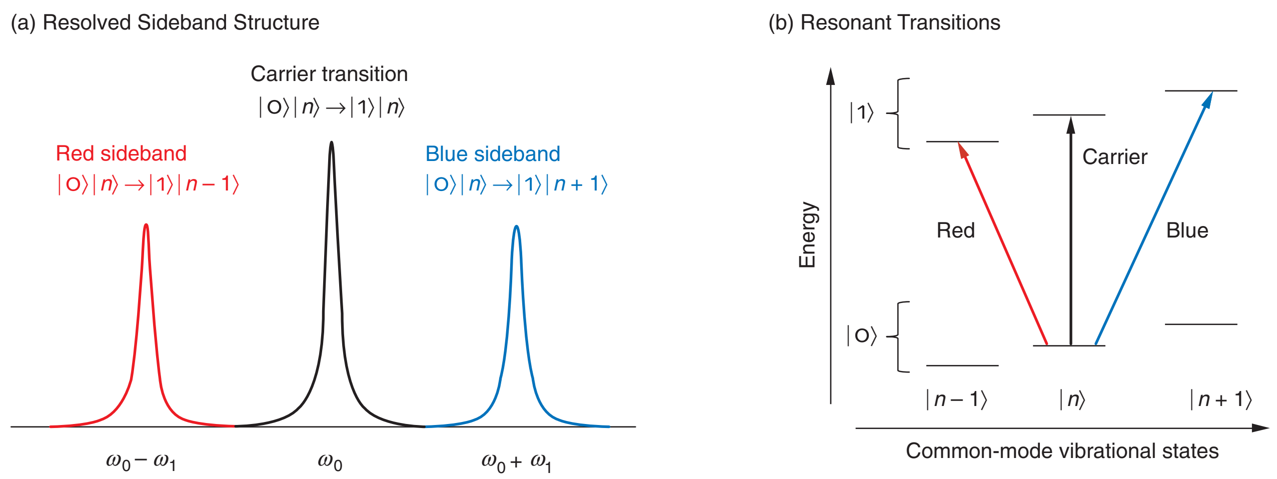


Figure 3. Vibrational Sideband Spectrum

(a) An ion trap naturally couples an ion's electronic excitations to its vibrational motion. Each electronic transition at resonant frequency ω_0, known as the carrier frequency, is therefore accompanied by other sideband transitions. We show the two sidebands closest in frequency to the carrier: the lower-energy red sideband at frequency ($\omega_0 - \omega_1$), and the higher-energy blue sideband at frequency ($\omega_0 + \omega_1$). A laser with a sufficiently narrow linewidth can interact with the ion via a specific sideband or the carrier. (b) Interacting with a particular qubit (ion) via a sideband transition will change the qubit's internal state and simultaneously the external state of all the qubits in the trap, either increasing the number of phonons in the common mode by one (excitation on the blue sideband) or decreasing the number by one (excitation on the red sideband).

While we can use laser pulses to interact with each qubit separately (and excite a qubit's electronic, or internal, degrees of freedom), we can also use another laser to excite the trap's vibrational modes and hence to interact with all qubits simultaneously. The latter process can be viewed as interacting with the qubits' external degrees of freedom. The state of a string of j qubits in the trap is therefore explicitly given as

$$|q_1, q_2, \dots q_j\rangle|n\rangle \quad . \tag{1}$$

The first ket, $|q_1, q_2, \dots q_j\rangle$, refers to the logical qubit states, with $q_j = 0$ or 1. The second ket, $|n\rangle$, refers to the common-mode vibrational state, and the value of n, say, 0, 1, 2, …, indicates the number of phonons in the common mode. (Although the string of qubits may initially be in another mode, we will restrict our attention to the common mode.) Thus, in the state

$$|q_1, q_2, \dots q_j\rangle|0\rangle \quad , \tag{2}$$

the ions are not vibrating because there are no phonons in the common mode. In the state

$$|q_1, q_2, \dots q_j\rangle|1\rangle \quad , \tag{3}$$

the common mode contains one phonon and all the ions are swaying in unison along the trap axis.

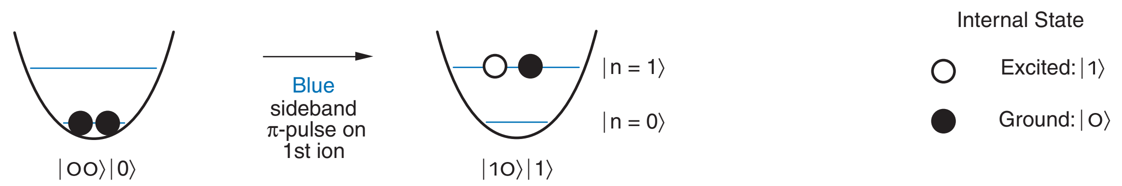
As mentioned earlier, the common mode is used as a "bus" that couples different ions together. To better understand this coupling, consider first that the frequency of the transition between the $|0\rangle$ state and the $|1\rangle$ state is ω_0, and that the common-mode vibrational frequency ω_1 is much lower than ω_0. Similar to the case of two coupled harmonic oscillators, the energy spectrum of the ion exhibits resonances not only at the "carrier" frequency ω_0, but also at the "sideband" frequencies $\omega_0 \pm \omega_1$ (see Figure 3). The resonance with the higher frequency is commonly known as the blue sideband; the one with the lower frequency, as the red sideband. For cold ions, the linewidth $\Delta\omega_0$ of the carrier transition is very narrow[3] and is less than the energy difference between the carrier and either sideband. Thus, the sidebands and the carrier can be resolved within the cold ion's frequency spectrum.

Now consider, for example, a procedure used to place two ions in an entangled state. Assuming that the ions are initially in the state $|00\rangle|0\rangle$, if we were to address the first ion with a π-pulse from a laser detuned to the blue sideband of the internal transition, both the internal and external states of that ion would be excited. Because an excitation in the common mode is felt by both ions, the result would be the two-qubit state $|10\rangle|1\rangle$. If, instead, we address the first qubit

---

[3] The metastable excited state has a very long lifetime, which leads to a narrow linewidth according to Heisenberg's uncertainty principle. To take an example from the Los Alamos experiment, a calcium ion excited to the $3^2D_{3/2}$ state will decay to the ground state only after an average delay of about 1 second, which results in a transition linewidth of about 1 hertz.

(a) Coupling Two Qubits through the Common Mode



(b) Entangling Two Qubits

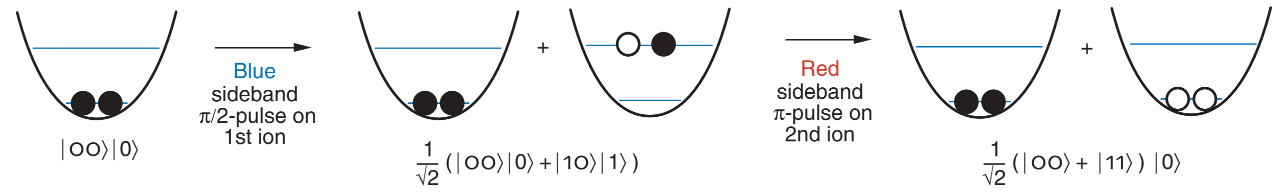

$|00\rangle|0\rangle$
Blue
sideband
$\pi/2$-pulse on
1st ion
+
$\frac{1}{\sqrt{2}}(|00\rangle|0\rangle + |10\rangle|1\rangle)$
Red
sideband
π-pulse on
2nd ion
+
$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)|0\rangle$

Figure 4. Using the Common Mode to Entangle Two Qubits

The vibrational state is indicated by the position of the ions on the rungs of a ladder in the harmonic potential well. In this diagram, the electronic ground state of an ion is indicated by a solid circle; the excited state, by an open circle. (a) Suppose two qubits are initialized to the state $|00\rangle|0\rangle$. Addressing the first qubit with a π-pulse from a laser tuned to the blue sideband will excite the ions to the state $|10\rangle|1\rangle$. The first ion is in its electronic excited state, while the second remains in its electronic ground state. Because the common mode affects all ions, both ions are excited to the $|n = 1\rangle$ vibrational state. (b) Two qubits can be entangled if we illuminate the first qubit with a $\pi/2$-pulse on the blue sideband. The ions are placed in a superposition of states: $1/\sqrt{2}(|00\rangle|0\rangle + |10\rangle|1\rangle)$. If the second ion is then illuminated by a π-pulse from a laser tuned to the red sideband, it can absorb the photon only if energy is available from the vibrational mode. Thus, ion 2 is excited only if ion 1 was excited; it remains in the ground state if ion 1 was in the ground state. The two-ion system therefore exhibits the strong correlation of observables, which according to Bohr, define the condition of entanglement. The end result of the operation is the entangled state $1/\sqrt{2}(|00\rangle + |11\rangle)|0\rangle$.

with a $\pi/2$-pulse (see Figure 4), both qubits are placed in a superposition of the two states, namely,

$$1/\sqrt{2}(|00\rangle|0\rangle + |10\rangle|1\rangle) \quad . \tag{4}$$

We then address the second ion (which is still in its ground state) with a π-pulse tuned to the red sideband. The laser energy is too low to excite directly the ground-to-excited-state electronic transition, but the transition still occurs if extra energy can be taken from the common mode. The end result is that all phonons have been removed from the quantum register at the end of the operation, and we create a two-qubit entangled state:

$$1/\sqrt{2}(|00\rangle + |11\rangle)|0\rangle \quad . \tag{5}$$

We can no longer describe the system as individual ions being in the ground or the excited state. The result of a measurement on one ion is strongly correlated to the status of the other ion. Notice that this procedure works equally well if one or more ions are placed in between the first and second ions because the excitation of the common mode is shared by all ions.

Besides defining the individual operations just described, Cirac and Zoller also spelled out in detail the steps needed to perform a "controlled-not" (**cnot**) gate. In such an operation, a "target qubit" flips its state only if a second qubit, the "control qubit," is originally set to its logical $|1\rangle$ value. Dave Wineland's group at NIST first implemented the **cnot** gate in an ion trap in 1995 (Monroe et al. 1995), albeit using only a single ion in the trap. (The two states of the control qubit were the two lowest-energy trap vibrational states.) Still, because any computation can be performed with a number of two-qubit **cnot** gates, together with some single-qubit gate operations, the realization of a **cnot** gate in a quantum computer is an important benchmark.

Readout. At the end of any quantum calculation, the individual qubits in the quantum register will be in defined states, which must be read out with high fidelity. A powerful readout tool makes use of the phenomenon of quantum jumps (Sauter et al. 1986, Bergquist et al. 1986). The readout method is easily understood when one examines the generic ion-level scheme shown in Figure 5. The ion has two states that we identify as the logical qubit states $|0\rangle$ and $|1\rangle$. But the ions used for quantum computation also have a short-lived excited state $|s\rangle$ that is accessible from one of the qubit states, the $|0\rangle$ state, by laser excitation. When the laser drives the $|0\rangle$ —> $|s\rangle$ transition for a long period, the ion fluoresces and emits a huge number of photons. If that transition is not accessible because the ion was in the $|1\rangle$ state, there will be no fluorescence. Detection of a fluorescence signal, therefore, tells us that the qubit is in the $|0\rangle$ state, and we observe the "jumps" of the ion between the $|0\rangle$ and the $|1\rangle$ state as distinct jumps in the intensity of the fluorescence.

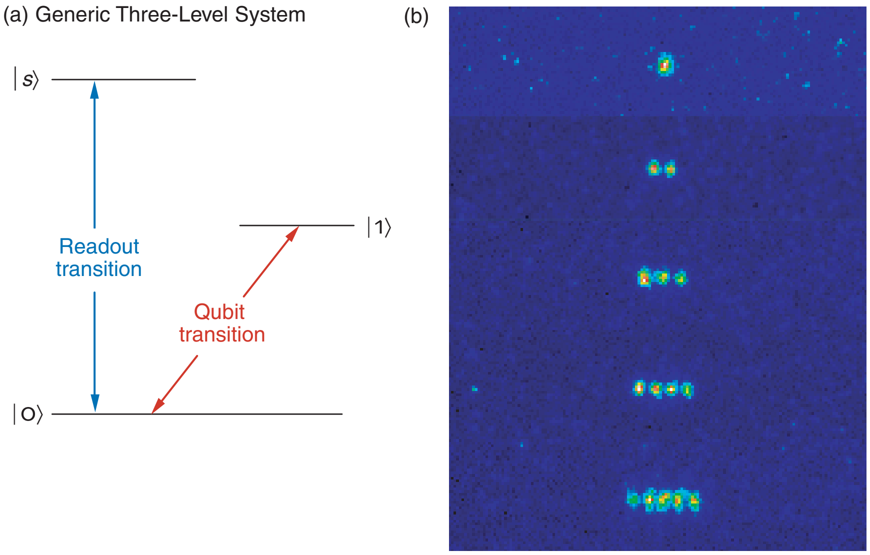


Figure 5. Readout Using Quantum Jumps

(a) A generic three-level scheme for ions in a trap is illustrated. The qubit states $|0\rangle$ and $|1\rangle$ are typically the ground state and a long-lived excited state, respectively. The state $|s\rangle$ is short lived and is coupled to the ground state. If the ion is in the ground state, a laser can drive the $|0\rangle \rightarrow |s\rangle$ transition many times per second, and the ion will fluoresce. If the ion "jumps" to the $|1\rangle$ state, there will be no fluorescence, so that the presence or absence of a large fluorescence signal reveals the state of the qubit. (Alternatively, one can use two long-lived ground-state hyperfine levels as qubits and construct a similar readout scheme.) (b) This composite image shows strings of calcium ions that were laser-cooled to near rest in the Los Alamos quantum computation ion trap. The spacing between the ions is approximately 30 µm. About 10^8 photons are absorbed and reemitted each second during the time the readout laser is irradiating the ion. That photon flux is easily detectable with a charge-coupled device (CCD) camera. The fluorescence is actually bright enough to be seen with the naked eye, except that for calcium, the readout transition is at 397 nm and is outside the range of sensitivity for the human eye.

This type of readout will destroy any quantum information contained in the qubit state and will yield a purely For example, suppose the ion is in an equal superposition of the states $|0\rangle$ and $|1\rangle$; then probing the ion once with the laser will not reveal the original state of the qubit. If we want to get a reading on the ratio of the two different states in a superposition, we will have to repeat the measurement multiple times and resort to a statistical description.

If we want to maintain the quantum character of the ion's state at the end of a particular calculation, we may resort to a different scheme. Consider an ion placed in a high-quality optical cavity, which is tuned to the resonance of the internal transition in the ion. If the ion is in the state $|1\rangle$, a photon is emitted into the cavity; if it is in the state $|0\rangle$, no photon is emitted. If the ion is in a superposition state, the photon field in the cavity will end up in a superposition between the states consisting of one photon in the cavity and no photon in the cavity. Thus, the quantum state of an ion or an atom can be transferred to a photon (Parkins and Kimble 1999, Mundt et al. 2002). This state could be transferred through optical fibers to a different trap and then transferred into another ion—and so, the quantum Internet is born!

The Los Alamos Ion-Trap Quantum Computing Experiment

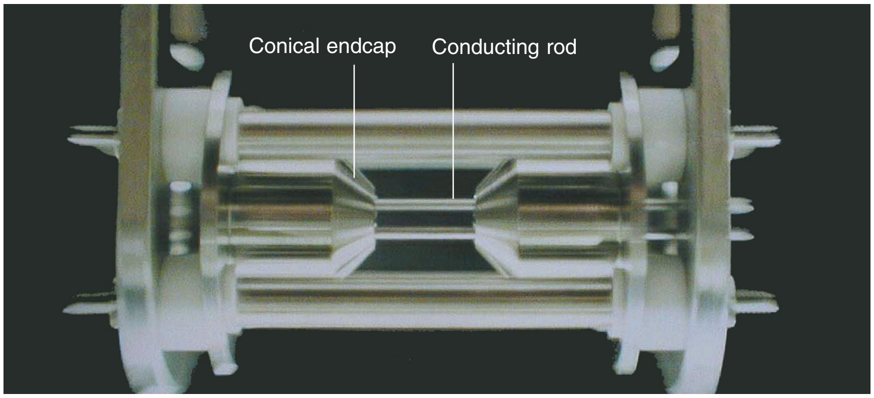
Currently, every implementation of ion-trap quantum computing uses qubits that are composed of two long-lived internal states of the trapped ions (the ground state and a metastable excited state, or two hyperfine sublevels of the ground state) and has the qubits communicating with each other through the trap's vibrational modes. Many different ion species can serve as qubits, and numerous qubit schemes are possible. While the previous section discussed ion-trap quantum computers in general terms, this section describes an experiment developed at Los Alamos, in which calcium ions were used.

We initially chose to use calcium for a number of reasons, including the following: All the wavelengths needed for cooling and manipulation are, at least in principle, accessible by relatively inexpensive diode lasers; the lifetime of the metastable state allows a reasonable number of coherent operations to be performed; and the calcium isotope of interest is most abundant and can easily be loaded into the trap. But the basic quantum computational schemes outlined earlier can be implemented with any element that displays an ionic-level structure similar to that of calcium, such as the other alkali-earth elements beryllium, magnesium, strontium, and barium. At this stage of experimentation, all alkali-earth ions are essentially interchangeable, and for mostly technical reasons, calcium has recently been replaced with strontium in our quantum computing experiment. (Some of that work is described in the article, "Quantum Information with Trapped Strontium Ions" on page 178.) In addition, other ions, such as mercury and ytterbium, also exhibit level schemes that are applicable to quantum computation, albeit with slightly different technical approaches. As ion-trap quantum computers become more sophisticated, the choice of ion species will become a larger issue.

Our trap is a linear Paul trap, about 1 centimeter in length and 1.7 millimeters in diameter, with a cylindrical geometry, as seen in Figure 6(a). We create the strong, radial confinement fields by applying a few hundred volts of rf amplitude at approximately 8 megahertz to two opposing rods. The remaining two rods are rf-grounded. The axial confinement, which prevents the ion from leaking out of the trap along the symmetry axis, is produced by a direct current of about 10 volts applied to each of the conical end caps. This combination of the rf and dc fields leads to an axial oscillation frequency ω_1 for the common mode of a few hundred kilohertz and a radial oscillation frequency of $\omega_r \approx 1$ megahertz.

Additional dc potentials can be applied to four support rods, which are not shown in Figure 6(a) but are located outside the actual trap electrodes. In this way, one can center the ion string on the electrical symmetry axis of the trap and thus minimize the amount of heating produced by micromotion.

(a) Trap Built at Los Alamos



(b) Partial Energy-Level Diagram for Calcium

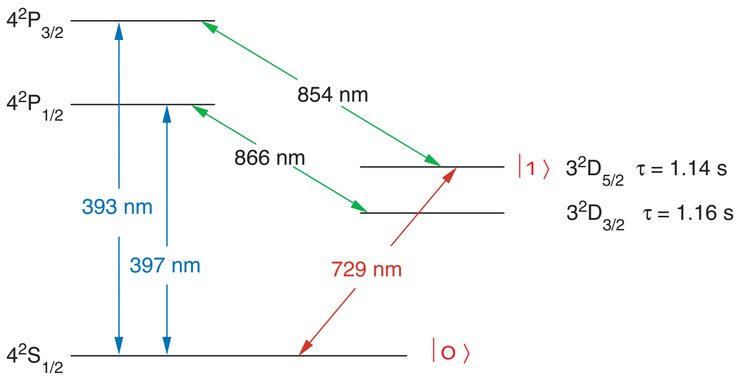


Figure 6. The Los Alamos Linear Paul Trap

(a) The trap built at Los Alamos for quantum computation is about 1 cm in length and 1.7 mm in diameter. An electric field of a few hundred volts oscillating at 8 MHz is applied to two of the conducting rods. The other two rods are RF grounded. About 10 V of a direct current is placed on the conical end caps. (b) This illustration shows a partial energy-level diagram for calcium (not to scale) and shows the wavelengths of some transitions relevant to our quantum computing scheme. The qubit transition is shown in red.

Figure 6(b) shows a schematic diagram of the internal-level structure of calcium ions and gives the wavelengths of the relevant transitions. (Any other alkali-like ion would have a similar structure.) The $4^2S_{1/2}$ ground state and the metastable $3^2D_{5/2}$ excited state are used to form the logical qubit states $|0\rangle$ and $|1\rangle$, respectively. The metastable excited state has a lifetime of about 1 second, which is long enough to allow an interesting number of computational steps to be performed before decoherence (resulting from spontaneous emission from the excited state) can destroy the internal state of the quantum register.

To load the ions into the trap, we cross a beam of calcium atoms that is produced by heating a small calcium-filled reservoir with a beam of electrons emitted by an "electron gun." (The electron gun is essentially identical to the one inside a computer monitor or a television screen.) These two beams are aligned so that they cross each other within the effective volume of the trap, that is, within the cylindrical volume that fits between the four rods and the two end caps. The atoms that are ionized by electron impact suddenly feel the confining forces of the electric fields and become trapped.

Cirac and Zoller (1995), as well as other authors, proposed initializing the computer in the state

$$|oo \ldots o\rangle|0\rangle \;; \qquad (6)$$

that is, all qubits are in their electronic and vibrational ground states. However, the temperature[4] of the newly trapped ions is very high, since their energy is given by a combination of the temperature of the calcium oven and the energy imparted to the ion by the electric field. (The latter energy varies, depending on where the ionization occurs.) In order to reach the initial state and then to perform quantum logic operations, the ions' temperature must be reduced to its lowest possible value. Cooling the ions takes place in two steps described in the next two sections.

Doppler Cooling of Calcium Ions. As its name suggests, this first cooling step makes use of the Doppler effect, whereby the relative motion between a source and an observer causes a change in the observed frequency of an acoustic or electromagnetic wave. For example, the sound of a siren on an ambulance or a police car changes its pitch depending on whether the vehicle moves toward you or away from you. Similarly, an ion or atom will absorb or emit photons of different frequencies (energies), depending on its motion relative to the light source. Although an ion in the trap is localized by electric fields and its average velocity is zero, the variation of the instantaneous velocity, as the ion jiggles back and forth due to thermal motion, causes the inherent emission and/or absorption profile of an ionic transition to become much broader than the natural linewidth of the transition (second-order Doppler broadening). For "hot" ions, the Doppler-broadened linewidth is typically much greater than the laser linewidth.

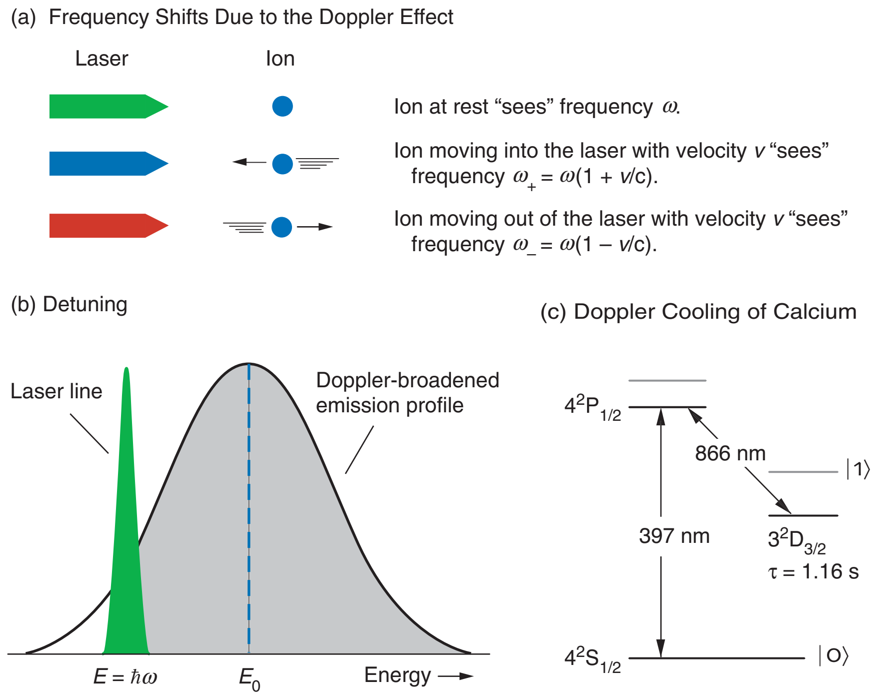

(a) Frequency Shifts Due to the Doppler Effect
Laser
Ion
Ion at rest "sees" frequency ω.
Ion moving into the laser with velocity v "sees" frequency $\omega_+ = \omega(1 + v/c)$.
Ion moving out of the laser with velocity v "sees" frequency $\omega_- = \omega(1 - v/c)$.
(b) Detuning
Laser line
Doppler-broadened emission profile
$E = \hbar\omega$
E_0
Energy
(c) Doppler Cooling of Calcium
$4^2P_{1/2}$
866 nm
$|1\rangle$
397 nm
$3^2D_{3/2}$
$\tau = 1.16$ s
$4^2S_{1/2}$
$|o\rangle$

Figure 7. Doppler Cooling of Ions

(a) When interacting with a laser of frequency ω, an ion at rest sees the native laser frequency. If the ion is moving, this frequency is shifted by the Doppler effect. An ion moving into the laser beam "sees" the laser frequency Doppler-shifted toward a higher frequency, ω_+, while the ion moving in the direction of the laser beam "sees" the frequency ω_-. (b) This first-order Doppler effect is eliminated in ion traps because the average velocity is zero. However, because of its thermal motion, the ion has a probability to absorb photons at any frequency within its Doppler-broadened absorption profile. Similarly, it has a probability to emit a photon over a range of frequencies within its emission profile. Suppose the laser is tuned below the ion's resonance frequency ω_0 so that $\omega < \omega_0$. When the ion moves into the laser beam, it will absorb a photon because it sees the laser frequency Doppler-shifted close to its resonance frequency ($\omega_+ \sim \omega_0$). The ion absorbs a laser photon of energy $E = \hbar\omega < \hbar\omega_0$, but on average it reemits a photon with higher energy (from the gray region). Because it loses energy during each cycle of absorption and emission, the ion cools rapidly to the limit of this method, which is imposed by the recoil energy the ion experiences upon reemission of the photon. For typical parameters of our trap, calcium will reach a vibrational level of approximately $|n = 10\rangle$ to $|n = 30\rangle$ at the end of the Doppler cooling. (c) The transitions used to Doppler-cool calcium ions are shown.

To implement Doppler cooling, we tune a laser to a frequency below the resonance frequency of a transition in the ion (Figure 7). Only when it is moving at a certain velocity toward the laser can the ion absorb these "off-resonance" photons, because only then does it "see" the laser frequency shift into resonance. However, as a result of its random jiggling, the ion has a probability to emit photons at any frequency within its Doppler-broadened emission line profile. One can easily see from Figure 7 that the ion has a greater probability to emit a photon with a higher frequency than the absorbed photon. On average, more energy is emitted than absorbed, which leads to a cooling of the ion.

---

[4] We often refer to ion temperature rather than energy because the ions show a distribution of energies over time.

For rapid cooling, a large number of photons must be absorbed and emitted, and therefore Doppler cooling is performed on a transition that can be cycled rapidly. We use the 397-nanometer transition from the $4^2S_{1/2}$ to the $4^2P_{1/2}$ state. The lifetime of the $4^2P_{1/2}$ state is about 10 nanoseconds, so the ion can absorb and reemit about 10^8 photons per second. Unfortunately, the $4^2P_{1/2}$ state has a chance of roughly 1 in 15 to decay into the metastable $3^2D_{3/2}$ state, which has a lifetime of about 1 second. The ion then takes so long to return to the ground state that it would be lost from the cooling cycle. To avoid this outcome and force ions to return from the $D_{3/2}$ level to the cooling cycle, we irradiate the ion with two lasers: the cooling laser at 397 nanometers and a "repump" laser at 866 nanometers.

Doppler cooling has its limits. Conservation of momentum guarantees that, after emitting a photon in one direction, the ion recoils in the opposite direction. Although this recoil energy is small, eventually it counteracts any cooling effects. For calcium ions, the Doppler limit is equivalent to a temperature of about 3 microkelvins. At that temperature, the kinetic energy of the ions is significantly less than the mutual Coulomb repulsion between ions. Essentially, they do not have enough kinetic energy to leap-frog each other, so the cold ions remain frozen in their relative locations and form a string. The photos in Figure 4 are examples of ion strings that were realized in our trap. At a 200-kilohertz common-mode frequency, the spacing between ions is about 30 micrometers.

Even at a temperature of 3 microkelvins, however, the ions have enough energy to occupy any of several vibrational modes, with many phonons per mode. (The specific distribution of states depends on the ions' temperature and the frequency of each mode.) Here, we will restrict our attention to the common mode. After Doppler cooling, the ions in the trap can typically occupy the states from $|n = 10\rangle$ to about $|n = 30\rangle$. Getting the qubits into the common-mode ground state ($|n = 0\rangle$), therefore, requires an additional cooling scheme.

Sideband Cooling of Calcium. We recall that ions in the trap can couple their internal degrees of freedom with their external motion, which leads to sidebands at $\omega_0 \pm \omega_1$, where ω_1 is the common-mode frequency, in the absorption spectrum. For cold ions with a minimal Doppler linewidth, these sidebands are resolvable from the carrier—see Figure 8(b). Thus, an ion can absorb photons not only at the carrier frequency ω_0 of their internal $|0\rangle \rightarrow |1\rangle$ transition but also on the upper and lower sidebands at the frequencies $\omega_0 \pm \omega_1$.

Assuming all ions are in the state $|0\rangle|n\rangle$, we can tune a laser with a suitably narrow linewidth to the red sideband—photon energy $[E_- = \hbar(\omega_0 - \omega_1]$—and excite one of the ions to the state $|1\rangle|n - 1\rangle$. In essence, energy is removed from the vibrational mode (the occupation number is reduced by one phonon) and is used to make up the deficit in photon energy. After its radiative lifetime, the ion can decay to one of three states: the state $|0\rangle|n - 2\rangle$, by emitting a photon with energy $[E_+ = \hbar(\omega_0 + \omega_1]$; the state $|0\rangle|n - 1\rangle$, by emitting a photon with energy $[E_0 = \hbar\omega_0]$; or a return to its initial state, by emitting a photon with energy $[E_- = \hbar(\omega_0 - \omega_1]$—see Figure 8(c). On average, the ion loses one vibrational photon of energy $E = \hbar\omega_1$ for each excitation–decay cycle. Because we started somewhere around $|n = 30\rangle$, we need about 30 cycles to bring the vibrational mode to its ground state(provided there are no competing effects that heat the ions while they are being cooled).

Unfortunately, the long lifetime of the $3^2D_{5/2}$ state is now a hindrance. In principle, we can scatter only one photon per second using this transition, which would render the process of cooling from $|n = 30\rangle$ to $|n = 1\rangle$ unacceptably slow. Heating processes —micromotion heating or simply radiative heating from other noise sources in the system—are much faster, and we would be unable to reach the desired starting point of all qubits being in the internal and external ground states.

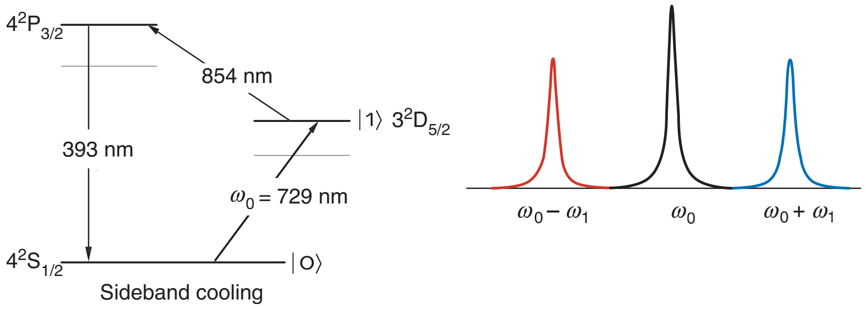
To speed things up, we artificially shorten the lifetime of the $3^2D_{5/2}$ state by introducing an alternate decay route via the $4^2P_{3/2}$ state (Marzoli et al. 1994). We irradiate the ion not only with a laser tuned to 729 nanometers (to drive the S–D transition), but also with a second laser tuned to 854 nanometers—see Figure 8(d). The second laser pumps the ion from the D- to the P-state, from which the ion rapidly returns to the ground state. By carefully choosing the amplitude of the 854-nanometer laser, we can design the effective lifetime of the $3^2D_{5/2}$ state according to our needs, and our calcium ion can jump down the ladder of harmonic-oscillator levels in just 3 to 30 milliseconds.

In a real system, the cooling power from the lasers will always be in competition with external heating processes. Although no clear theoretical explanation of these processes has emerged, many possibilities have been discussed in the literature, and the relevant scaling laws with trap parameters have been developed (James 1998b). Typical candidates—besides micromotion heating, which can be avoided by carefully tuning the trap voltages—are fluctuating contact potentials on the trap electrodes (originating from insulating deposits on the electrodes), which have a frequency component at the trap's resonant frequency.

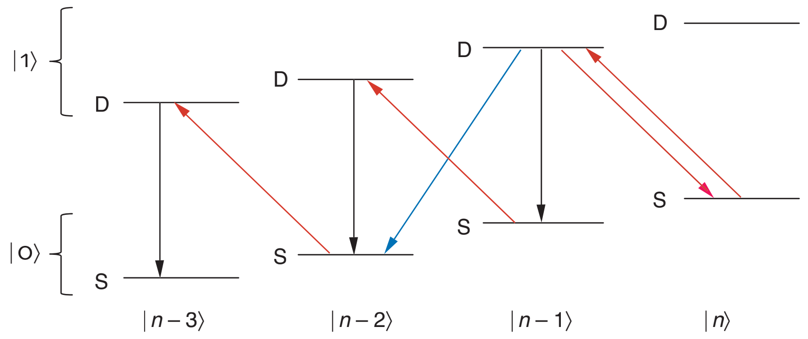
In the absence of a proper theoretical description of ion heating, we can

(a) Sideband Cooling Transitions

(b) Resolved Sideband Structure



(c) Stepping to Lower Vibrational States



(d) Increasing the Cooling Rate

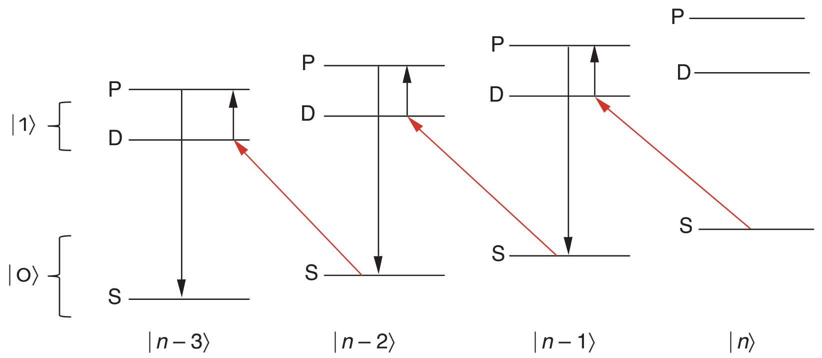


Figure 8. Sideband Cooling

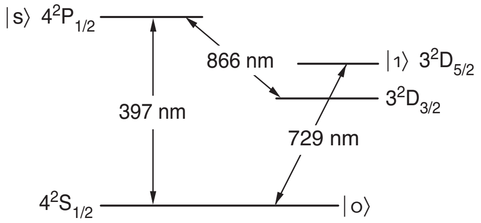
(a) This partial energy-level diagram shows the transitions we use for sideband cooling of calcium ions. (b) When the linewidth of the carrier transition (frequency ω_0) is very narrow and the Doppler broadening is minimal, the ion's vibrational sidebands can be resolved. (c) The figure shows several vibrational levels for the $|0\rangle \rightarrow |1\rangle$ carrier transition. If a single ion is initially in the state $|0\rangle|n\rangle$, then illuminating the ion with a laser tuned to the red sideband will excite the ion to the state $|1\rangle|n-1\rangle$. The latter state will decay to $|0\rangle|n-2\rangle$ or $|0\rangle|n-1\rangle$, or it will go back to $|0\rangle|n\rangle$. On average, the number of phonons in the mode decreases by 1 after each excitation/emission. (d) The lifetime of the upper level may be artificially shortened if that level is coupled to an auxiliary one with a higher decay rate. The faster decay will increase the cooling rate.

turn to empirical data accumulated from a number of different experiments. The adopted procedure is to cool the ions to as low a temperature as possible and then turn off the lasers responsible for the cooling. After a variable delay time, we measure the ion's temperature using sideband spectroscopy. Quentin Turchette and coworkers (2000) conducted the most complete study of this type when they looked at heating effects in traps of different sizes. The separate traps had also undergone different preparation "rituals." The studies suggest a strong dependence on trap size, that is, on the distance between the ions and the trap electrodes. When the studies are combined with observations made by Rainer Blatt's group at the University of Insbruck, one is led to believe that "bigger is better." But Ralph deVoe with IBM has recently reported that hardly any heating was observed over a short period in a miniaturized trap.

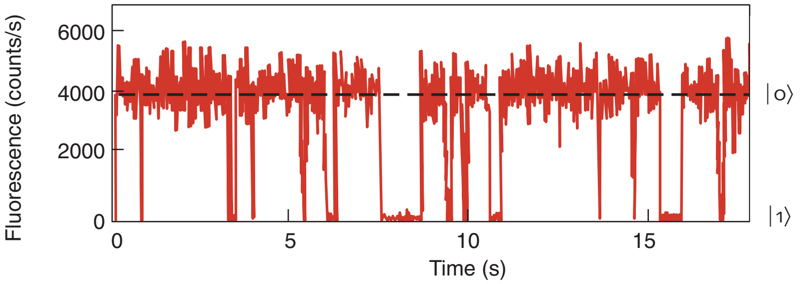
Clearly, we have much to learn before we can understand the heating of ions in rf traps. The comforting thought is that, in all cases, the time scale for heating from the ion's ground state can be kept long, compared with the time required for a reasonable number of quantum manipulations. Furthermore, heating times are typically longer than times for other decoherence processes.

Readout of the Calcium Ion. We use the $|s\rangle = 4^2P_{1/2}$ excited state in calcium for readout, the same state that is used for Doppler cooling. As discussed earlier, this state has a lifetime of only 10 nanoseconds and is accessed from the ground state by a laser tuned to 397 nanometers. An ion in the $|0\rangle$ state will absorb and reemit about 10^8 photons per second when the laser drives the $|0\rangle \rightarrow |s\rangle$ transition. (Because the $4^2P_{1/2}$ state can also decay to the long-lived $3^2D_{3/2}$ state, we simultaneously irradiate the ion with a laser tuned to

(a) Calcium Readout Transitions



(b) Quantum Jumps, Single Ion



(c) Quantum Jumps, Two Ions

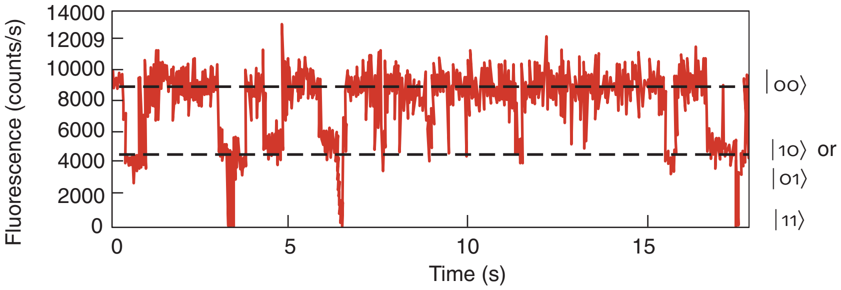


Figure 9. Readout of Qubits
(a) Shown here are the readout transitions for calcium. (b) For this readout experiment, a single ion interacts with two lasers: a low-intensity laser that drives the qubit transition $|0\rangle \rightarrow |1\rangle$ and a second laser that drives the readout transition $|0\rangle \rightarrow |s\rangle$. The fluorescence signal from that transition, nominally around 4,000 counts per second, is recorded with a simple rate meter. When the qubit is in the $|0\rangle$ state, we can drive the readout transition, but if the ion occupies the state $|1\rangle$, the fluorescence disappears. We can distinguish between the $|0\rangle$ and $|1\rangle$ states with nearly 100% fidelity. (c) The state of two ions can also be distinguished. No count corresponds to the state $|1, 1\rangle$; 8000 to 10,000 counts per second correspond to the state $|0, 0\rangle$; 4000 counts per second, to either $|01\rangle$ or $|10\rangle$. (In the last case, our experimental setup does not allow us to distinguish between the two states.)

866 nanometers to return the ion to the $4^2P_{1/2}$ state.) Even with a modest photon-collection efficiency of about 10^{-4}, which is due to experimental considerations (we cannot bring a lens too close to the ions without blocking access to the trap), we can easily detect the photons scattering from the ion with a charge-coupled device (CCD) camera.

In Figure 9, we show a sample trace of the detected photon counts for a single ion in the trap. The fluorescence signal is nominally about 10^4 counts per second. We randomly excite the ion with a laser tuned to 729 nanometers, and each time it "jumps" from the $|0\rangle$ state to the $|1\rangle$ state, the signal disappears. Figure 9 also shows the fluorescence from a set of two ions. The different levels of intensity are for both ions being excited (no fluorescence), for one of the two ions being excited (intermediate fluorescence), and finally, for both ions being in the ground state (full fluorescence). Although it is easy to distinguish among these cases, determining which of the two ions is in the ground state for the intermediate fluorescence level is difficult. We must look at the ions individually, by focusing the laser on one ion at a time, and then convert to the single-ion measurement.

Ferdinand Schmidt-Kaler and his colleagues from the Innsbruck group have used this readout technique with three ions, which were spaced at about 6 micrometers from each other in the trap. They cooled the ions to the $|000\rangle|n\rangle$ state, and all three were emitting photons on the readout transition. The scientists then pointed a sharply focused laser at 729 nanometers onto one of the ions and placed it in the $|1\rangle$ state (the dark state). The measured crosstalk among neighboring ions was less than 1 percent, so the state of the chosen qubit could be determined with about 99 percent fidelity (Nägerl et al. 1999).

Important Developments

A *Popular Mechanics* article from 1949 stated, "Where a calculator on the ENIAC (electronic numerical integrator and calculator) is equipped with 18,000 vacuum tubes and weighs 30 tons, computers in the future may have only 1000 tubes and weigh only one and a half tons." That observation did not turn out to be entirely correct. How could anyone have foreseen the development of transistors and integrated solid-state circuitry or the remarkable parallel developments that have culminated in today's supercomputers?

We are still in the "vacuum-tube" era of quantum computation, and if asked two years ago about the future of ion-trap-based quantum computers,

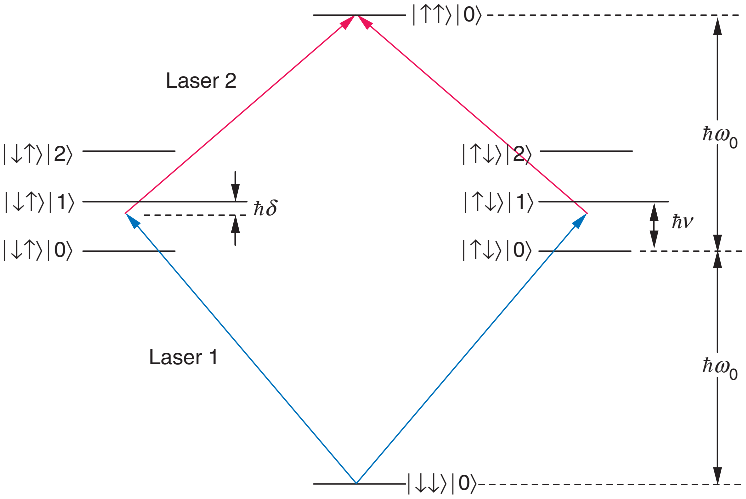

$|\uparrow\uparrow\rangle|0\rangle$
Laser 2
$|\downarrow\uparrow\rangle|2\rangle$
$|\downarrow\uparrow\rangle|1\rangle$
$\hbar\delta$
$|\downarrow\uparrow\rangle|0\rangle$
$|\uparrow\downarrow\rangle|2\rangle$
$|\uparrow\downarrow\rangle|1\rangle$
$\hbar\nu$
$|\uparrow\downarrow\rangle|0\rangle$
$\hbar\omega_0$
Laser 1
$\hbar\omega_0$
$|\downarrow\downarrow\rangle|0\rangle$

Figure 10. Four-Particle Entanglement

The figure shows the relevant energy levels and transition frequencies used to create deterministic multiparticle entanglement. A two-ion scheme is illustrated. The $|\uparrow\uparrow\rangle|0\rangle$ excited state has an energy of $2E_0 = 2\hbar\omega_0$. The $|\uparrow\downarrow\rangle|1\rangle$ and $|\downarrow\uparrow\rangle|1\rangle$ excited states, in which the internal state of one of the ions is excited and both ions go into a vibrational excited state, has an energy $E_1 = \hbar(\omega_0 + \nu)$. Lasers tuned to energies $E_1 + \delta$ and $E_1 - \delta$, where δ is a predetermined laser detuning, can directly excite the ions to the $|\uparrow\uparrow\rangle|0\rangle$ state. Pulsing the two lasers for a time $t = \pi/(2\Omega)$, where Ω is an effective Rabi frequency, will place the ions in the entangled state $|\Psi_2\rangle = 1/\sqrt{2}(|\uparrow\uparrow\rangle - i|\downarrow\downarrow\rangle)$. The scheme can be generalized to any number of ions and has been used to create entangled states of up to four ions.

(Figure reproduced with permission from *Nature*.)

I would have been hesitant to promise much. I may have argued that the systems we were looking at were mere demonstrations, designed to help us understand the fundamental physics issues behind qubits and that the prospects for scaling these devices up to a larger number of qubits were doubtful. Even today I could argue that, while the computing scheme of Cirac and Zoller is in principle scalable (Hughes et al. 1996), it has yet to be realized with two qubits.

However, because much has happened in the ensuing two years, included here are descriptions of just a few of the many important developments that have put the ion-trap quantum computer back on the track for scalable technologies. Similar to the transition from vacuum tubes to solid-state devices (even if not quite as fundamental), these developments do not invalidate any of the previous achievements and underlying principles but are unpredicted and significant enhancements of available technology.

Four-State Entanglement. To take full advantage of the power of quantum computation, we need to generate entanglement between an arbitrary number of qubits. But generating any entangled state is difficult. In the case of photons, entanglement is achieved by means of a statistical process. Many pairs of photons are created by a method known as parametric down-conversion, whereby a high-energy photon, after entering a special type of crystal, has a certain probability to convert into two photons, each with half the initial energy. In a few cases, two photons emerge in an entangled state. (See the article "Quantum State Entanglement" on page 52.) We can typically produce about 1000 entangled pairs per second, but if we look for entanglement of three or even four photons, the likelihood of finding such a state becomes unacceptably small for practical purposes (30 per second for 3 photons and a few per year for four photons).

Thus, quantum computing took a leap forward when the NIST team in Boulder demonstrated that it could produce an entangled state of up to four ions "on demand" (Sackett et al. 2000). Based on a proposal by Anders Mølmer and Klaus Sørenson (1999) from the University of Aarhus in Denmark, the NIST team around Chris Monroe and David Wineland demonstrated the feasibility of entanglement of two and four ions in a deterministic way. With a single-pulse operation of two lasers, the desired state could be produced with a high degree of certainty.

To understand the technique, consider two spin-half particles confined in a harmonic well and coupled by vibrational degrees of freedom. (The spin description is equivalent to our previous picture of two internal states in an ion.) The NIST group used the two ground-state hyperfine levels of $^9\text{Be}^+$ ions as an effective spin-half system, with $|\downarrow\rangle = |F = 2, m_F = -2\rangle$ and $|\uparrow\rangle = |F = 1, m_F = -1\rangle$. The energy levels of the system are shown in Figure 10, where $\hbar\omega_0$ is the internal energy splitting of each particle and ν is the oscillation frequency of the particular collective mode of the particles in the trap.

The group used standard laser-cooling and optical-pumping techniques to prepare the particles in their spin-down internal state and in the ground state of their collective motion: $|\Psi\rangle = |\downarrow\downarrow\rangle|0\rangle$. Laser pulses at $\omega_0 + (\nu - \delta)$ and $\omega_0 - (\nu - \delta)$, where δ is the detuning from the resonance (refer to Figure 10), then drive the two-step transition from $|\downarrow\downarrow\rangle|0\rangle$ to $|\uparrow\uparrow\rangle|0\rangle$. If the detuning δ is sufficiently large, the intermediate states $|\uparrow\downarrow\rangle|1\rangle$

and $|\downarrow\uparrow\rangle|1\rangle$ are negligibly occupied, and no motional excitation occurs in the process. Applying the laser fields for a time $t = \pi/(2\Omega)$, where Ω is the Rabi oscillation frequency for the transition, results in the final wave function

$$|\Psi_2\rangle = 1/\sqrt{2}\,(|\uparrow\uparrow\rangle - i\,|\downarrow\downarrow\rangle)\ , \qquad (7)$$

which is the desired maximally entangled state.

It turns out that this process is entirely scalable for an even number of N qubits and can generate the N-particle entangled state

$$|\Psi_N\rangle = 1/\sqrt{2} \times (|\uparrow\uparrow\ldots\uparrow\rangle - i^{N+1}|\downarrow\downarrow\ldots\downarrow\rangle)\ . \qquad (8)$$

If N is an odd number, the state $|\Psi_N\rangle$ can still be produced, provided one rotates each qubit independently. The NIST scientists have used this method with two and four ions in the trap, but they also caution that the experimentally realized states $|\Psi_2\rangle$ and $|\Psi_4\rangle$ are not fully entangled. Each state shows some degree of decoherence. Although the amount of decoherence in $|\Psi_4\rangle$ was more than what could be tolerated for quantum computing, the achievement of reliably creating a four-particle entangled state on demand cannot be underestimated.

In a later development, the NIST group showed that the maximally entangled states of a pair of trapped $^9\mathrm{Be}^+$ ions could be used as a decoherence-free subspace for protecting one qubit of information against dephasing errors (Kielpinski et al. 2001). The decoherence-free subspace, also called a noiseless subsystem, is spanned by the two orthogonal states

$$|\Psi_+\rangle = 1/\sqrt{2}\,(|\downarrow\uparrow\rangle + i\,|\uparrow\downarrow\rangle)\ , \text{ and}$$

$$|\Psi_-\rangle = 1/\sqrt{2}\,(\,|\downarrow\uparrow\rangle - i\,|\uparrow\downarrow\rangle)\ . \qquad (9)$$

These states serve as the logical qubit for storing information. It is easy to see that all superpositions of these maximally entangled states are invariant under transformations that apply the phase change $|\uparrow\rangle \rightarrow e^{i\varsigma}|\uparrow\rangle$ simultaneously to both ions. This so-called collective dephasing is thought to be a major source of decoherence for trapped ions.

In the NIST experiment, an arbitrary state of one qubit was encoded in the decoherence-free subspace of two ions:

$$\alpha|\uparrow\rangle + \beta|\downarrow\rangle \rightarrow \alpha|\Psi_+\rangle + \beta|\Psi_-\rangle\ . \qquad (10)$$

The encoded information was subjected to engineered dephasing errors or ambient errors, and then the encoding procedure was reversed to recover the original information. The data showed unequivocally that the noiseless subsystem protects the information from collective dephasing errors for a time up to ten times longer than the typical decoherence time and that collective dephasing is indeed a major source of errors in ion traps. One could imagine that this type of robust storage might enable the operation of a quantum computer constructed from an array of ion traps as opposed to a single trap. (For an introduction to the theory of noiseless subsystems, see the article "Introduction to Quantum Error Correction" on page 188. A nuclear magnetic resonance experiment demonstrating noiseless subsystems is presented in the article "Realizing a Noiseless Subsystem in an NMR Quantum Information Processor" on page 260.)

Broadband Cooling. The second important recent result is the selective enhancement of the probability of cooling ions by electromagnetically induced transparency (EIT). The scheme of Cirac and Zoller has the qubits coupled together by means of the common vibrational mode, in which all ions oscillate back and forth in unison along the trap axis. However, even two trapped ions have an extra degree of freedom in the axial motion, namely, the breathing mode, in which ions on opposite sides of the string move 180° out of phase (refer to Figure 2). Each additional ion opens up three more vibrational modes to the ion string. Every mode of frequency ν can be assigned an average quantum number n_ν.

The initial scheme of Cirac and Zoller requires a mode to have $n_\nu = 0$ in order to be used for computational operations. For small numbers of ions, we reach this state by the standard sideband-cooling methods discussed earlier. As seen in Figure 11(a), the ion has a number of transition possibilities: Excitation on the lower sideband will cool the ion, excitation on the upper sideband will cause heating, and transitions on the carrier will cause diffusion. In sideband cooling, we use an ultranarrow laser and excite only the lower sideband so that $|n\rangle \rightarrow |n-1\rangle$.

For a large number of qubits, however, the sheer number of higher modes makes it technically difficult, if not impossible, to use standard sideband-cooling methods. Not only would we have to identify and excite the lower-sideband transitions for each and every mode, but the spectrum becomes so "dense" that the upper sidebands of a neighboring internal transition can overlap the lower sidebands of another. Cooling one mode could actually heat another. Furthermore, the "overhead" needed to control and cool these modes is daunting: large numbers of laser pulses, constant retuning of the lasers from one mode to the next, and tight control of the qubit register throughout the cooling stage.

For efficient (and simultaneous) cooling of more than one mode, broadband cooling would be required, even though that would seemingly exacerbate the problem of unwanted excitation. But recent work by Blatt's group at the University of Innsbruck

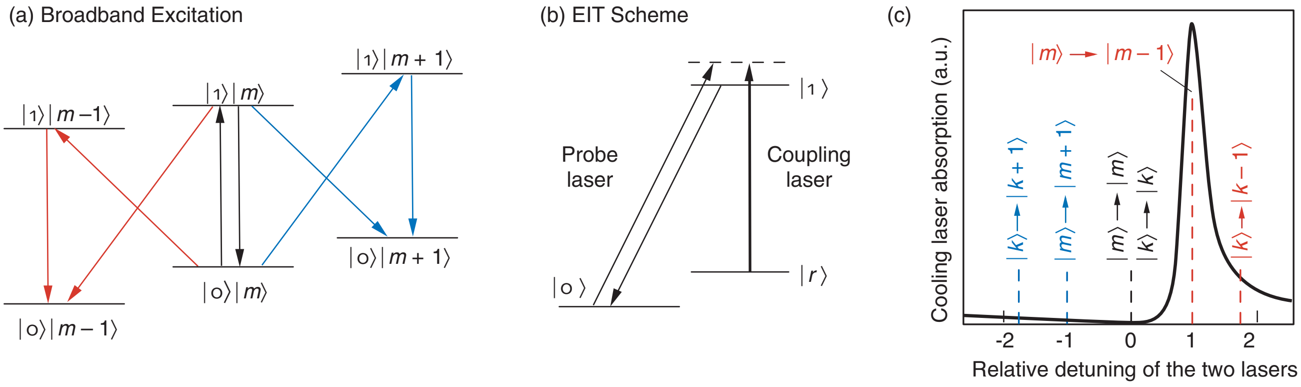

(a) Broadband Excitation
$|1\rangle|m-1\rangle$
$|1\rangle|m\rangle$
$|1\rangle|m+1\rangle$
$|0\rangle|m-1\rangle$
$|0\rangle|m\rangle$
$|0\rangle|m+1\rangle$
(b) EIT Scheme
Probe laser
Coupling laser
$|1\rangle$
$|0\rangle$
$|r\rangle$
(c)
Cooling laser absorption (a.u.)
$|m\rangle \rightarrow |m-1\rangle$
$|k\rangle \rightarrow |k+1\rangle$
$|m\rangle \rightarrow |m+1\rangle$
$|m\rangle \rightarrow |m\rangle$
$|k\rangle \rightarrow |k\rangle$
$|k\rangle \rightarrow |k-1\rangle$
-2
-1
0
1
2
Relative detuning of the two lasers

Figure 11. EIT Cooling

Sideband cooling of a multi-ion string that is accessing many excited vibrational modes is very difficult in that the sideband structure becomes dense and complicated. EIT cooling permits broadband cooling of several vibrational modes, $|m\rangle$, $|k\rangle$, ... , simultaneously. (a) When a broadband probe laser is applied to the $|0\rangle|m\rangle \rightarrow |1\rangle|m\rangle$ transition, both cooling (red) and heating (blue) transitions can occur. (b) When a second coupling laser excites the $|r\rangle \rightarrow |1\rangle$ transition, the ion's absorption profile becomes modified. Proper choice of laser detuning (to the dashed state) suppresses heating transitions. This result is evident in figure (c), where the solid line gives the absorption profile for the EIT scheme. For proper tuning of the lasers, the absorption strength for the transition $|m\rangle \rightarrow |m\rangle$ is zero and a strong asymmetry between $|m\rangle \rightarrow |m+1\rangle$ and $|m\rangle \rightarrow |m-1\rangle$ transitions is introduced. This asymmetry in absorption between the blue and the red sideband also holds for higher-frequency vibrational modes ($|k\rangle \rightarrow |k \pm 1\rangle$), allowing simultaneous cooling of several different modes with one broadband laser. **[Figure was adapted from Schmidt-Kaler (2001) with permission from the authors.]**

may make broadband cooling possible (Morigi et al. 2000, Roos et al. 2000). The group adopted the EIT technique to selectively enhance the probability of exciting cooling transitions rather than heating transitions in the ion.

The necessary asymmetry between lower and higher sidebands can be achieved as follows: Consider a three-level system with two lower levels and one shared excited state—see Figure 11(b). Using a strong coupling laser between one of the ground states and the upper state creates so-called light shifts (that is, shifted energy levels, as seen by another probe laser). For a detuning of the coupling laser above the resonance, a probe laser sees an absorption profile that shows zero absorption for a detuning equal to the coupling laser, the so-called Fano profile—see Figure 11(c). Therefore, such a probe would be transparent for that exact detuning—the EIT phenomenon. In order to obtain optimum cooling using these EIT resonances, the detunings are chosen such that the carrier transition is exactly located at the EIT resonance (that is, it is not excited at all because of that quantum interference), and the maximum absorption is chosen to be around the lower sideband frequency.

Because the absorption profile generated in this manner is fairly wide (much wider than the natural width of the transition used for traditional sideband cooling), the asymmetry between heating and cooling transitions exists for many modes. Several different modes can be cooled simultaneously with a single operation. This technique reduces the overhead for laser-cooling of multi-ion strings and also eases the requirements for laser stability, which are very strict for standard sideband cooling.

To show that EIT cooling can simultaneously cool vibrational modes with significantly different frequencies of oscillation, the Innsbruck group chose to cool the axial mode and the radial mode of a single ion confined in a three-dimensional Paul trap at 3.3 megahertz and 1.6 megahertz, respectively (Schmidt-Kaler et al. 2001). In a linear trap, the nearby modes ("spectator" modes) are not used for the computation directly; they are coupled to and may affect the common mode. The group achieved ground-state populations of 73 percent for the axial and 58 percent for the radial mode. Although this result is certainly not as satisfactory as that achieved by sideband cooling (because of the smaller absorption asymmetries), it is certainly sufficient for cooling (and thus suppressing) those modes. The EIT method promises the possibility of cooling all spectator modes of a multiqubit quantum register with a single operation. That would allow the more elaborate (individual) sideband cooling scheme to be used on only the mode needed for calculations.

Outlook

Many systems have been proposed in the last several years as potential candidates for becoming quantum computers, including laser-cooled trapped ions (Cirac and Zoller 1995), nuclear magnetic resonance (Gershenfeld and Chuang 1997, Cory et al. 1998), cavity quantum electrodynamics (Ye et al. 1999), and more recently, superconducting devices, quantum dots, and silicon-based solid-state devices.

From the preliminary experiments performed by several groups worldwide, it is apparent that the existing ion traps are adequate to hold and manipulate small numbers of qubits. Although five to ten qubits hardly a computer make, these numbers are large enough to make the technology well worth pursuing. Ion traps will be a potent tool for exploring, for example, the possibility of creating entangled states of large numbers of qubits. Investigations of the type described here will help us identify the relevant physics issues that must be addressed to achieve computational gains.

We should also expect that many of the technologies now being pursued for quantum computation will be superseded by even more promising ideas. One such idea is to scale up to a larger number of qubits by multiplexing several ion traps with a specific trap that contains a few qubits acting as the central processor. After implementing part of a quantum algorithm, the qubits would be shuffled into one of several storage traps, thus allowing new qubits to be loaded into the processor. Recent work also suggests that we could transfer the internal quantum states of a string of ions in a trap to a set of photons in a high-finesse cavity. The quantum information could then be transferred through optical fibers into a second cavity and fed back into an ion string in a different trap. Developments like this will surely continue to happen and will allow us to explore quantum computation well beyond the current state of the art.

As we get closer to realizing a small quantum processor, the “time scales” of a particular system become more relevant. In general, the hierarchy of time scales present in an ion-trap quantum computer is very promising. Manipulations on quantum registers can be done in microseconds, while disturbances by the environment have been shown to be avoidable for milliseconds. The inherent decoherence time of the quantum state is longer still, for it is limited by the lifetime of the upper qubit state, which is about 1 second in calcium. The decoherence time can be increased even more by an appropriate choice of ions (for example, ytterbium) or by stable ground-state hyperfine levels used as logical qubit states.

It is important to point out that despite the revolutionary advances in computers during the last 50 years, the fundamental principle of computation has not changed. Today’s fastest supercomputer operates according to the same rules as the ENIAC. Quantum computation, however, represents a paradigm shift in information processing. Although a future quantum computer may not look anything like our current ion trap, the experience and knowledge we gain now will be of fundamental importance to our understanding this new paradigm of computing.

For some researchers, building a quantum computer to break secure codes is an important, and certainly challenging, goal. But for me and most of my colleagues, performing experiments that Erwin Schrödinger and Albert Einstein only dreamed of and thus gaining a deep understanding of this “inconceivable” quantum world are far larger rewards. Perhaps we will encounter some failure of conventional quantum mechanics, or perhaps the problems of decoherence will forever keep the quantum realm out of our classical grasp. In any event, the future will be exciting for both quantum physics and computation. ■

Acknowledgments

The work described in this article is the result of a close and fruitful collaboration among numerous experimental scientists over several years. I wish to thank them all for their help in the laboratory, as well as for the many intellectual discussions that helped me understand in depth the fundamental principles involved. Over the years, I enjoyed working with Daphna Enzer, John Gomez, Mark Gulley, Paul Kwiat, Steve Lamoreaux, Glen Peterson, Vern Sandberg, Martin Schauer, Dale Tupa, and Justin Torgerson on developing capabilities for quantum computation with trapped ions. In addition, I had the privilege to travel to many groups pursuing this dream and learned much of what I know today about the ion-trap quantum processor from many interactions with my friends and colleagues Rainer Blatt, Ferdinand Schmidt-Kaler, Dietrich Leibfried, Christoph Nägerl, and many others at the University of Innsbruck in Austria; with Andrew Steane, David Lucas, and Derek Stacey from the Clarendon Laboratory in Oxford, the United Kingdom; and last but not least, with all the members in Dave Wineland’s group at NIST, in Boulder (too numerous to name here individually), who took me on challenging excursions into the quantum world of trapped ions, as well as into the all too classical world of mountain biking.

Further Reading

Bergquist, J. C., R. G. Hulet, W. M. Itano, and D. J. Wineland. 1986. Observation of Quantum Jumps in a Single Atom. *Phys. Rev. Lett.* **57**: 1699.

Cirac, J. I., and P. Zoller. 1995. Quantum Computations with Cold Trapped Ions. *Phys. Rev. Lett.* **74**: 4091.

Cory, D. G., M. D. Price, and T. F. Havel. 1998. Nuclear Magnetic Resonance Spectroscopy: An Experimentally Accessible Paradigm for Quantum Computing. *Physica D* **120** (1–2): 82.

Dawson, P. H., ed. 1976. *Quadrupole Mass Spectrometry and Its Applications*. Chaps. II and III. Amsterdam: Elsevier Scientific Publishing Co.

Dehmelt;, H. G. 1967. Radiofrequency Spectroscopy of Stored Ions, Part I. *Adv. At. Mol. Phys.* **3**: 53.

———. 1969. Radiofrequency Spectroscopy of Stored Ions, Part II. *Adv. At. Mol. Phys.* **5**: 109.

———. 1981. Coherent Spectroscopy on Single Atomic System at Rest in Free Space 2. *J. Phys.* (*Paris*) **42**: 299.

———. 1988. A Single Atomic Particle Forever Floating at Rest in Free Space: New Value for Electron Radius. *Physica Scripta*. **T22**: 102.

Enzer, D. G., M. M. Schauer, J. J. Gomez, M. S. Gulley, M. H. Holzscheiter, P. G. Kwiat et al. 2000. Observation of Power-Law Scaling for Phase Transitions in Linear Trapped Ion Crystals. *Phys. Rev. Lett.* **85** (12): 2466.

Gershenfeld, N. A., and I. L Chuang. 1997. Bulk Spin-Resonance Quantum Computation. *Science* **275**: 350.

Huang, X.-P., J. J. Bollinger, T. B. Mitchell, and W. M. Itano. 1998. Phase-Locked Rotation of Crystallized Non-Neutral Plasmas by Rotating Electric Fields. *Phys. Rev. Lett.* **80**: 73.

Hughes, R. J., D. F. V. James, E. H. Knill, R. Laflamme, and G. F. Petschek. 1996. Decoherence Bounds on a Quantum Computation with Trapped Ions. *Phys. Rev. Lett.* **77**: 3240.

Hughes, R. J., D. F. V. James, J. J. Gomez, M. S. Gulley, M. H. Holzscheiter, P. G. Kwiat, et al. 1998. The Los Alamos Trapped Ion Quantum Computer Experiment. *Fortschr. Phys.* **46** (4–5): 329.

James, D. F. V. 1998a. Quantum Dynamics of Cold Trapped Ions with Application to Quantum Computation. *Appl. Phys. B* **66** (2): 181.

———. 1998b. Theory of Heating of the Quantum Ground State of Trapped Ions. *Pys. Rev. Lett.* **81**: 317.

Kielpinski, D., V. Meyer, M. A. Rowe, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland. 2001. A Decoherence-Free Quantum Memory Using Trapped Ions. *Science* **291**: 1013.

Marzoli, I., J. I. Cirac, R. Blatt, and P. Zoller. 1994. Laser Cooling of Trapped Three-Level Ions: Designing Two-Level Systems for Sideband Cooling. *Phys. Rev. A* **49**: 2771.

Mølmer, K., and A. Sørensen. 1999. Multiparticle Entanglement of Hot Trapped Ions. *Phys. Rev. Lett.* **82** (9): 1835.

Monroe, C., D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland. 1995. Demonstration of a Fundamental Quantum Logic Gate. *Phys. Rev. Lett.* **75**: 4714.

Morigi, G., J. Eschner, and C. H. Keitel. 2000. Ground State Laser Cooling Using Electromagnetically Induced Transparency. *Phys. Rev. Lett.* **85** (21): 4458.

Mundt, A. B., A. Kreuter, C. Becher, D. Leibfried, J. Eschner, F. Schmidt-Kaler, and R. Blatt. 2002. Coupling a Single Atomic Quantum Bit to a High Finesse Optical Cavity. http://eprints.lanl.gov. quant-ph/0202112.

Nägerl, H. C., D. Liebfried, H. Rhode, G. Thalhammer, J. Eschnerr, F. Schmidt-Kalerr, and R. Blatt. 1999. Laser Addressing of Individual Ions in a Linear Ion Trap. *Phys. Rev. A* **60** (1): 145.

Neuhauser, W., M. Hohenstatt, P. E. Toschek, and H. Dehmelt. 1980. Localized Visible Ba^+ Mono-Ion Oscillator. *Phys. Rev. A* **22**: 1137.

Parkins, A. S., and H. J. Kimble. 1999. Quantum State Transfer between Motion and Light. *J. Opt. B*, *Quantum Semiclassical Opt.* **1** (4): 496.

Paul, W., H. P. Reinhard, U. von Zahn. 1958. Electrical Mass Filters as Mass Spectrometers and Isotope Filters. *Z. Phys.* **152**: 143.

Raizen, M. G., J. M. Gilligan, J. C. Bergquist, W. M. Itano, and D. J. Wineland. 1992. Ionic Crystals in a Linear Paul Trap. *Phys. Rev. A* **45**: 6493.

Roos, C. F., D. Leibfried, A. Mundt, F. Schmidt-Kaler, J. Eschner, and R. Blatt. 2000. Experimental Demonstration of Ground State Laser Cooling with Electromagnetically Induced Transparency. *Phys. Rev. Lett.* **85** (26): 5547.

Sackett, C. A., D. Kielpinski, B. E. King, C. Langer, V. Meyer, C. J. Myatt, et al. 2000. Experimental Entanglement of Four Particles. *Nature* **404**: 256.

Sauter, Th., W. Neuhauser, R. Blatt, and P. E. Toschek. 1986. Observation of Quantum Jumps. *Phys. Rev. Lett.* **57**: 1696.

Schmidt-Kaler, F., J. Eschner, G. Morigi, C. F. Roos, D. Liebfried, A. Mundt, and R. Blatt. 2001. Laser Cooling with Electromagnetically Induced Transparency: Application to Trapped Samples of Ions or Neutral Atoms. *Appl. Phys. B* **73** (8): 807.

Steane, A., C. F. Roos, D. Stevens, A. Mundt, D. Leibfried, F. Schmidt-Kaler, and R. Blatt. 2000. Speed of Ion-Trap Quantum-Information Processors. *Phys. Rev. A* **62**: 042305.

Tanoudji, C. C., B. Diu, and F. Laloe. 1977. *Quantum Mechanics*, Vol. I, Chap. IV. B. 3a, p. 403. New York: John Wiley and Sons.

Turchette, Q. A., D. Kielpinski, B. E. King, D. Leibfried, D. M. Meekhof, C. J. Myatt, et al. 2000. Heating of Trapped Ions from the Quantum Ground State. *Phys. Rev. A* **61**: 063418.

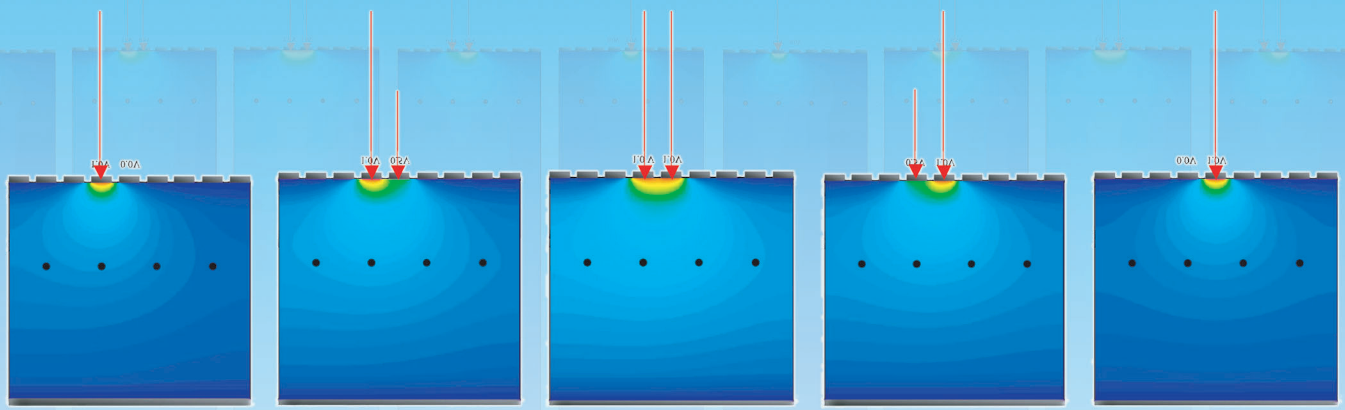
Walther, H. 1991. In *Light Induced Kinetic Effects on Atoms*, *Ions*, *and Molecules*, p. 261. Edited by L. Moi et al. (Pisa, Italy: ETS Editrice).

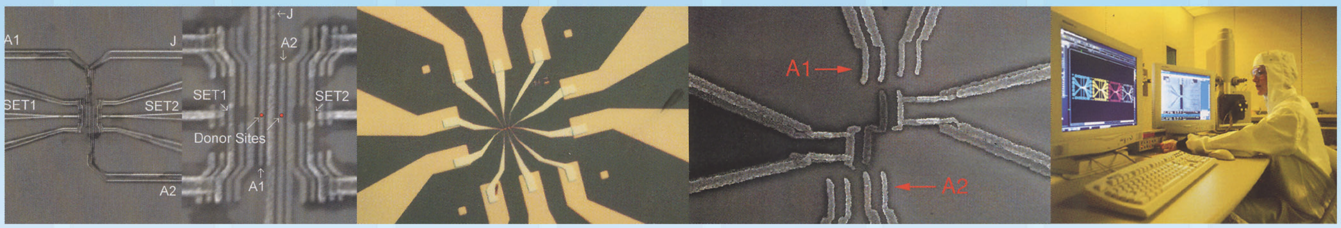
———. 1994. Atoms in Cavities and Traps. *Adv. At. Mol. Opt. Phys.* **32**: 379.

Wineland, D., P. Ekstrom, and H. Dehmelt. 1973. Monoelectron Oscillator. *Phys. Rev. Lett.* **31**: 1279.

Ye, J., D. W. Vernooy, and H. J. Kimble. 1999. Trapping of Single Atoms in Cavity QED. *Phys. Rev. Lett.* **83** (24): 4987.

Michael Holzscheiter received his M.S. and Ph.D. degrees in physics from the University of Mainz in Germany, where he studied electrons trapped in Penning traps. As a postdoctoral researcher at Texas A&M University, Michael used trapped ions to study collisional processes of astrophysical relevance. Later, as an assistant professor at Texas A&M, he participated in a collaboration with Los Alamos National Laboratory on trapping antiprotons. In 1986, he joined Los Alamos and pioneered the dynamic trapping of high-energy particles in Penning traps. He became principal investigator of the Los Alamos antiproton experiment, which was installed at CERN in 1992. On the basis of his technique's success, he formed an international collaboration, ATHENA, to create antihydrogen atoms at rest for ultrahigh precision studies of the symmetries between matter and antimatter and served as spokesman for this collaboration from 1995 through 1999. Back in Los Alamos, Michael applied his expertise in ion traps toward a Los Alamos trapped-ion quantum computer, for which he designed the radio-frequency quadrupole trap and helped build and operate the first-generation experiment until the fall of 2001. His research interest continues in the application of the ion-trap technology to a wide variety of physics problems.





Toward a Silicon-Based Nuclear-Spin Quantum Computer

Developing the technology for a scalable, solid-state quantum computer

Robert G. Clark, P. Chris Hammel, Andrew Dzurak, Alexander Hamilton,
Lloyd Hollenberg, David Jamieson, and Christopher Pakes
as told to Jay Schecker

One of the major challenges in quantum computing is to identify a system that can be scaled up to the number of qubits needed to execute nontrivial quantum algorithms. Peter Shor's algorithm for finding the prime factors of numbers used in public-encryption systems (numbers that typically consist of more than a hundred digits) would likely require a quantum computer with several thousand qubits. Depending on the error correction scheme appropriate to the particular computer, the required number could be much larger. Although ion-trap or nuclear-magnetic-resonance (NMR) quantum "computers" containing a few (less than 10) qubits have been successfully demonstrated, it is not clear that those systems can be scaled up.

Solid-state quantum computers may be more likely candidates. As a result, a number of solid-state schemes have been proposed, most of which fall into two categories: The physical qubits are spin states of individual electrons or nuclei, or they are charge or phase states of superconducting structures.

A particularly promising scheme is the silicon-based nuclear-spin computer, proposed a few years ago by Bruce Kane (1998), then of the University of New South Wales in Sydney, Australia, and now of the University of Maryland in College Park, Maryland. Shown in Figure 1, the Kane computer architecture consists of a linear array of phosphorus atoms embedded beneath the surface of a pure silicon wafer. Each phosphorous atom serves as a qubit, and the linear array forms the computer's quantum register.[1] The entire wafer is placed in a strong, static magnetic field $\mathbf{B}_0$ at sub-Kelvin temperatures, and alignment of the phosphorous atom's nuclear spin as parallel or antiparallel to $\mathbf{B}_0$ corresponds to the $|0\rangle$ and $|1\rangle$ states of the qubit, respectively. (Throughout this article, we will follow the convention of Kane and use the notation |0⟩, |1⟩ to designate both qubit and nuclear spin states. We will use arrows, |↓⟩ or |↑⟩, to designate electron spin states.)

---

[1] Two-dimensional qubit arrays are also possible but require complex electrode geometries and additional insulators.

In order to execute a quantum algorithm, we need to rotate individual qubits in Hilbert space and couple two qubits together. We accomplish both operations with an array of gate electrodes[2] that lies on top of the wafer but is isolated from the pure silicon by a thin insulating layer of silicon dioxide (SiO_2). Referring to Figure 1, notice that each A-gate sits precisely above a phosphorous atom and each J-gate lies between adjacent atoms. As discussed later, a small positive voltage applied to the A-gate gives independent control of the qubit directly under the gate, while voltage applied to the J-gate allows coupling two qubits together through an electron-mediated interaction.

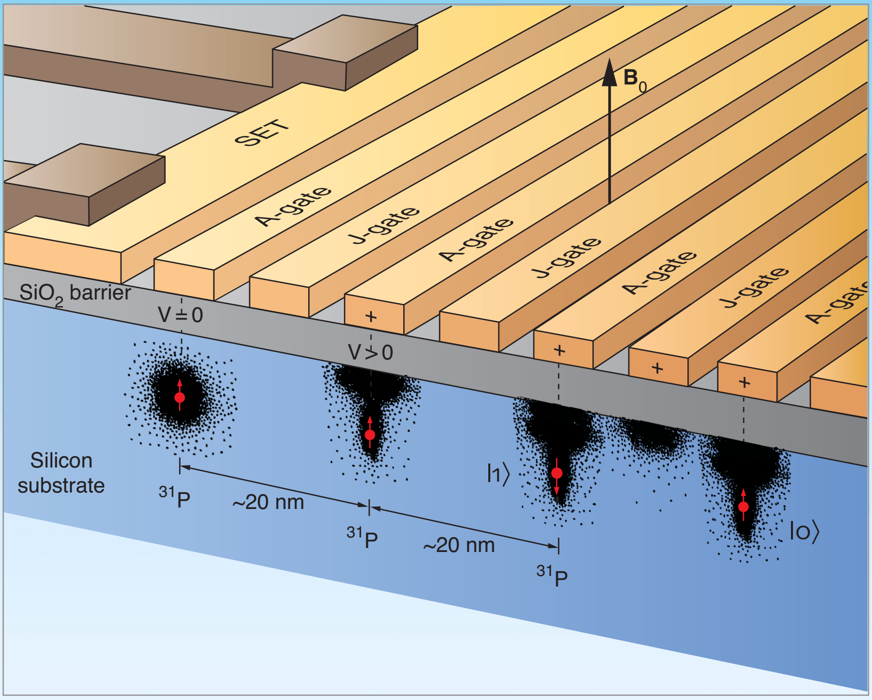


Figure 1. Schematic of the Silicon-Based Quantum Computer

The architecture of Kane's solid-state quantum computer has a linear array of phosphorous donor atoms buried into a pure silicon wafer, with an interdonor spacing of about 20 nm. In the presence of a large magnetic field and at sub-Kelvin temperatures, the nuclear spins of the donor atoms can be aligned either parallel or antiparallel with the field, corresponding to the |0⟩ and |1⟩ qubit states, respectively. An array of metal gates lies on the surface of the wafer, isolated from the silicon by a barrier layer of SiO_2. The A-gates lie directly above the donor atoms and enable individual control of single qubits. The J-gates lie between the donors and regulate an electron-mediated coupling between adjacent nuclear spins, thus allowing for two-qubit operations. Readout of the qubit is performed with either a single electron transistor, as shown, or with a magnetic-resonance force microscope (MRFM, not shown). The electron clouds shown here are schematic—the actual situation is more complicated.

The Center for Quantum Computer Technology (CQCT), headquartered in Sydney, Australia, and Los Alamos National Laboratory are trying to fabricate Kane's silicon-based quantum computer. Although we can call upon the technology, techniques, and collective experience of the huge silicon semiconductor industry, building the computer is still a daunting enterprise. We need to manipulate individual phosphorus atoms and place them precisely within a defect-free, isotopically pure silicon matrix. We need to create metal gates on the nanoscale, lay them within a few atoms of each other, and then ensure that each gate is aligned properly over the buried qubits. Simply put, the ability to do this level of nanofabrication does not exist at this time.

Employing a "see-what-works-best" strategy, we have initiated parallel research approaches for nearly every fabrication stage. If one approach fails, we still have a backup. Our current focus is on developing a prototype two-qubit device. By stretching many technologies beyond their heretofore-assumed limits, we have come tantalizingly close to achieving that goal. In the sections that follow, we describe the computer and some critical technologies in greater detail, and we also outline our progress in building a prototype.

Design Features of the Silicon-Based Computer

In our solid-state architecture, individual phosphorus atoms are embedded in a sea of silicon. These two elements were chosen for several reasons, the first and foremost being that phosphorous is the standard dopant for conventional silicon-based semiconductor devices and there is tremendous working knowledge of phosphorus and silicon from the conventional computing industry.

The second reason stems from the need to control the spin environment. We require our qubits to have nuclear spin $I = 1/2$, but we also want the surrounding environment to be spin free. Otherwise, unwanted spin-spin interactions between the qubit and any nearby nuclear spins would compromise the coherent states needed for quantum computation. The only stable phosphorous isotope, phosphorous-31, is a spin-1/2 nucleus, so nature has automatically satisfied our qubit criterion. Creating a spin-free environment, however, is a little more difficult. Natural silicon contains a mixture of three isotopes: silicon-28, -29, and -30. Whereas the even-numbered isotopes are spin free, silicon-29 has a spin of $I = 1/2$. As a result, we estimate that to do quantum computation, we will need to reduce the silicon-29 content in our silicon wafer to about one part in 10^5. Those stringent isotopic purity levels can be reached with current technology.

---

[2] In this context, a "gate," like a transistor gate, is a device used for controlling charge motion. It does not refer to a logical operation such as a **cnot** gate.

Finally, the nuclear spin of a phosphorous atom in a silicon host is stable. One way to infer the stability is to examine the spin-lattice relaxation time T_1, which characterizes the time it takes for a spin system with a net alignment (a net magnetization) to return to its thermally equilibrated magnetization. At temperatures of about 1 kelvin, the nuclear-spin relaxation time $T_{1,n}$ (where the subscript "n" is for the nucleus) for phosphorus in silicon is expected to be longer than 10 hours (Feher 1959).

The nuclear spin qubits, however, interact with the environment through their donor electrons; as a result, the electron spin stability is also important, particularly for qubit readout (see later discussion). The electron-spin relaxation time $T_{1,e}$ (where the subscript "e" is for the electron) is approximately 1 hour at about 1 kelvin (Honig and Stupp 1960). The phase decoherence time of the electron spin, as measured by the spin-spin relaxation time $T_{2,e}$, is shorter still. Although never measured for an isolated electron system such as our qubit scheme, the $T_{2,e}$ for an ensemble of electrons was measured to be approximately 0.5 millisecond (Gordon and Bowers 1958). A recent theoretical study, appropriate for a single phosphorus donor atom in silicon, indicates a $T_{2,e}$ of the order of 1 second (Mozyrsky et al. 2002). This value for $T_{2,e}$ should be long enough for us to perform a quantum computation and read out the result.

The Spin Hamiltonian and Single-Qubit Operations. To understand the physics underlying the operation of the silicon-based computer, recall that phosphorus has one more electron in its outer electron shell than silicon. When it is placed into a silicon crystal lattice, phosphorus fulfills its role as a surrogate silicon atom and still has one electron left over. At very low temperatures, that "donor" electron remains bound—albeit rather loosely—to the phosphorus nucleus. The electron "talks" to the nucleus primarily through the charge (Coulomb) interaction and to a lesser degree through the hyperfine interaction, which is between the electron spin and the nuclear spin.

We exploit the hyperfine interaction to individually address single qubits. The effective low-energy, low-temperature Hamiltonian describing the spin interactions for the electron spin and the nuclear spin of an atom in the presence of a static magnetic field $\mathbf{B}_0$ is given by

$$H = \mu_B \mathbf{B}_0 \sigma_z^e - g_n \mu_n \mathbf{B}_0 \sigma_z^n + A\sigma^e \cdot \sigma^n \ , \qquad (1)$$

where σ_z^e and σ_z^n are Pauli spin matrices, μ_B and μ_n are the Bohr and nuclear magnetons, respectively, and g_n is the nuclear g-factor. The hyperfine interaction is described by the term $A\sigma^e \cdot \sigma^n$.

For large values of $\mathbf{B}_0$, the Hamiltonian in Equation (1) leads to a set of energy levels that correspond to the four hyperfine levels, $|0\downarrow\rangle$, $|1\downarrow\rangle$, $|0\uparrow\rangle$, and $|1\uparrow\rangle$. At the sub-Kelvin operating temperature of the computer, however, the electron spins are completely spin-polarized in the lower-energy state $|\downarrow\rangle$. Thus, for one-qubit operations, we may ignore the electron spin polarization to a good approximation and consider only the two nuclear states $|0\rangle$ and $|1\rangle$ (Goan and Milburn 2000). The energy difference between those states is

$$\Delta E_0 = \hbar\omega_0 \cong 2g_n \mu_n \mathbf{B}_0 + 2A + 2A^2/\mu_B \mathbf{B}_0 \ , \qquad (2)$$

where ω_0 is called the nuclear resonance frequency. The resonance frequency, which is typically in the radio-frequency (rf) range, is equal to the Larmor frequency, or the rate at which the nuclear spins precess about the magnetic-field lines.

Suppose $\mathbf{B}_0$ is aligned along the z-axis, and the nuclear spin is initially in the $|0\rangle$ state. If we subject the spins to a secondary magnetic field that is oscillating in the x-direction at the nuclear resonance frequency, that is, a field $\mathbf{B}_1 = B_1 \cos(\omega_0 t)\, \hat{x}$, then the nuclear spins will begin to rotate toward the $(-z)$-axis, or from the parallel to the antiparallel alignment (see the box "Spin Manipulation with Magnetic Resonance" on page 288). The spin rotation is equivalent to rotating a qubit in Hilbert space from the $|0\rangle$ state to some superposition of the $|0\rangle$ and $|1\rangle$ states.

As described, the $\mathbf{B}_1$ field will affect all spins simultaneously. To address a particular spin, we use the A-gate directly above it and modify that donor atom's hyperfine coupling. The parameter A in Equation (1) is proportional to the magnitude of the electron probability density at the site of the nucleus, $\Psi_e(0)$:

$$A = 8/3\pi \ \mu_B \, g_n \, \mu_n |\Psi_e(0)|^2 \ . \qquad (3)$$

As seen in Figure 2, placing a positive voltage on the A-gate above the phosphorous atom attracts the atom's electron cloud toward the surface and away from the nucleus, thereby reducing the magnitude of $\Psi_e(0)$. The hyperfine energy levels of that one atom change slightly, and the resonance frequency needed to rotate the nuclear spin is reduced from, say, ω_0 to ω_-. If the frequency of the $\mathbf{B}_1$ field is set to ω_-, that is, $\mathbf{B}_1 = B_1 \cos(\omega_- t)\,\hat{x}$, then only the spin directly under the A-gate will be in resonance and will begin to rotate. Removing the voltage on the A-gate halts the rotation.

A one-qubit gate operation is therefore implemented if the silicon wafer is subjected to a transverse oscillating B-field of frequency ω_- and if the A-gate above a qubit is pulsed for a defined period. Throughout the duration of the pulse, the qubit is in resonance with the secondary magnetic field and rotates through some angle in Hilbert space. When the voltage is removed at the end of the pulse, the qubit is left in the desired superposition of the $|0\rangle$ and $|1\rangle$ states.

Two-Qubit Operations. To select adjacent pairs of qubits for two-qubit operations, we apply a positive voltage to the J-gate between them. As seen in Figure 3(a), this procedure causes the electron wave functions of the two donor atoms to overlap, and the electron spins couple together through the electron-spin exchange interaction. Because each electron is coupled to its nucleus through the hyperfine interaction, turning on the electron-spin exchange interaction also couples the nuclear spins together.

The coupled nuclear-electron spin system is fairly complex, but we can gain insight into it by looking at the effective Hamiltonian for the system:

$$H_{\text{coupled}} = H^1 + H^2 + J\sigma^{1e}\cdot\sigma^{2e} \quad . \qquad (4)$$

This Hamiltonian is valid at energy scales that are small compared with the electron-binding energies of the donor atoms. The first two terms correspond to the hyperfine Hamiltonian—Equation (1)—of each donor, respectively, and the last term accounts for the spin exchange

Continued on page 290

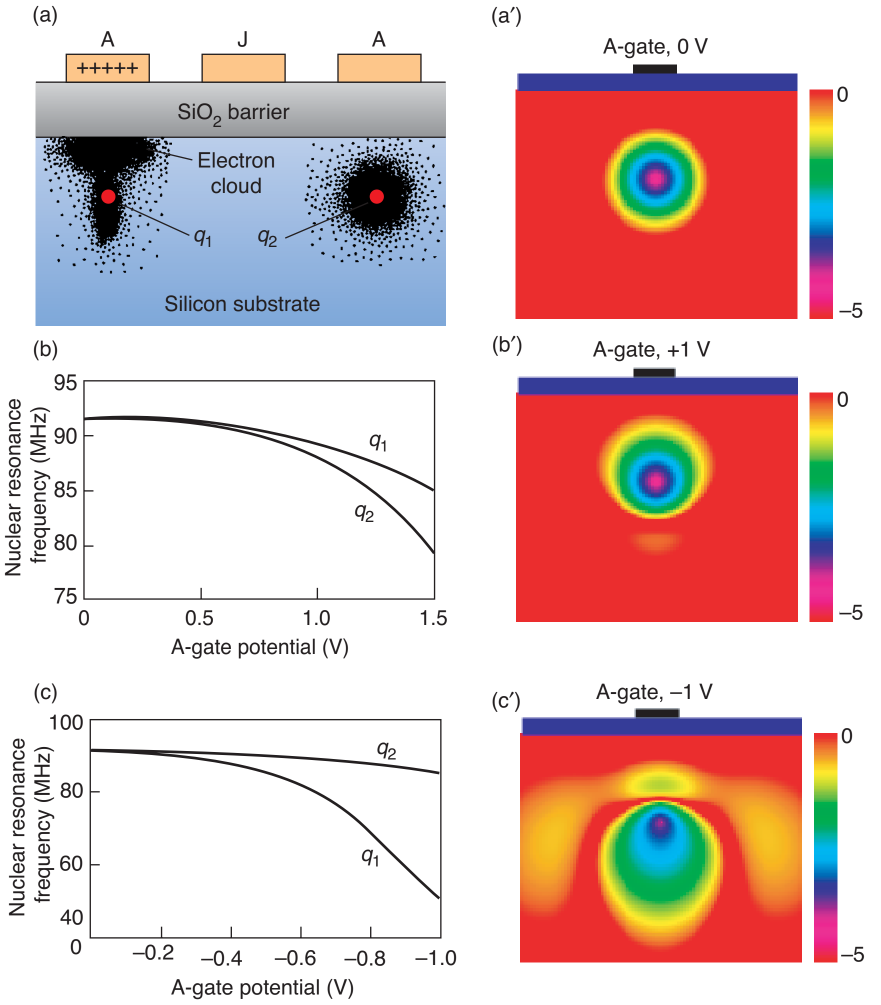
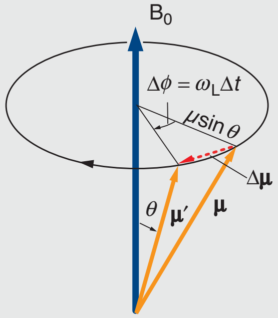


Figure 2. A-Gate Control of One Qubit

We use magnetic resonance techniques to rotate nuclear spins and place them in arbitrary superpositions of $|0\rangle$ and $|1\rangle$ qubit states. (a) In this cartoon, a small voltage is applied to the A-gate directly above a qubit. The donor electron moves away from the ^{31}P nucleus. (a′) This plot of the electron probability density surrounding a donor atom with no voltage on the A-gate was obtained by solving the Schrödinger equation nonperturbatively in an isotropic effective-mass hydrogenic basis. The plot is a cross section through the nucleus, with the color variations on a logarithmic scale. The atom is buried 20 nm below the Si/SiO$_2$ interface. (b) The graph shows the variation of the nuclear transition frequency as a function of A-gate voltage. (b′) The color plot shows the electron probability density. A positive voltage on the A-gate pulls the electron away from the nucleus, thus reducing the hyperfine coupling—the parameter *A* in Equation (1) in the text. In a B-field of about 2 T, the resonance frequency of a phosphorous nucleus q_1 is $\nu_0 = \omega_0/2\pi \approx 93$ MHz. With a gate voltage of 1 V, the resonance frequency of q_1 reduces to about $\nu_- = \omega_-/2\pi \approx 90$ MHz, while a neighboring nucleus q_2 is in resonance at about 87 MHz. (The proximity of the oxide barrier has a fairly large effect on the qubits, and the positive gate voltage affects q_2 more than q_1.) Subjecting the silicon wafer to a transverse, oscillatory magnetic field of frequency ν_- would cause only q_1 to respond. (c)–(c′) Initial calculations indicate that the electron probability density is more responsive to a negative gate bias, which results in better frequency discrimination between adjacent qubits.

Spin Manipulation with Magnetic Resonance



$$
\begin{aligned}
|\boldsymbol{\tau}| = |\boldsymbol{\mu} \times \mathbf{B}_0| &= \mu B_0 \sin\theta \\
\boldsymbol{\tau} \cong \Delta\mathbf{J}/\Delta t &= \gamma^{-1}\Delta\boldsymbol{\mu}/\Delta t \\
|\boldsymbol{\tau}| &= \gamma^{-1}(\mu \sin\theta\, \omega_L \Delta t)/\Delta t \\
&= \gamma^{-1}\mu \sin\theta\, \omega_L \\
\gamma^{-1}\mu \sin\theta\, \omega_L &= \mu B_0 \sin\theta \\
\Rightarrow \omega_L &= \gamma B_0
\end{aligned}
$$

Figure A. Larmor Precession
Magnetic moments precess around magnetic field lines at the Larmor precession frequency ω_L, which is derived above.

Magnetic resonance is the traditional technique for detecting and manipulating any particle, such as an electron, atom, or nucleus, that has a magnetic moment $\boldsymbol{\mu}$. The manipulation is controlled by a combination of static and oscillating magnetic fields. Classically, a particle's magnetic moment is proportional to its angular momentum **J** through the relation

$$\boldsymbol{\mu} = q/2m\, \mathbf{J} \ , \tag{1}$$

where q is the charge of the particle and m is its mass. Remarkably, a similar relation holds true in quantum mechanics, although we must also take into account the particle's intrinsic spin angular momentum. In general, we can write

$$\boldsymbol{\mu} = \gamma \mathbf{J} \ , \tag{2}$$

where the parameter γ is known as the gyromagnetic ratio. It is related to the constants in equation (1) by a dimensionless constant known as the g-factor,

$$\gamma = \mathrm{g}\ (q/2m) \ . \tag{3}$$

The magnitude and sign of the g-factor depend on the specific atom or nucleus, but are always approximately 1.

In the classical picture of a randomly oriented moment in a magnetic field $\mathbf{B}_0 = B_0\hat{z}$, the moment would like to lower its energy by aligning itself parallel to the applied field. But the magnetic field can only produce a torque on the moment, $\tau = \boldsymbol{\mu} \times \mathbf{B}_0$. Because the torque is directed perpendicular to the plane defined by the field, the moment does not align with the field, but like a spinning gyroscope that resists the force of gravity, precesses around the magnetic field line. By using the fact that the torque is equal to the rate of change of the angular momentum, we can derive the angular precession frequency of the moment (see Figure A):

$$\omega_L = \gamma B_0 \ , \tag{4}$$

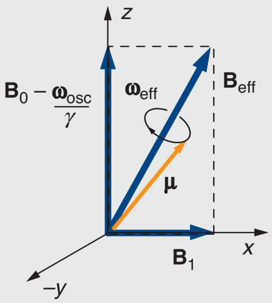

z
$\mathbf{B}_0 - \frac{\boldsymbol{\omega}_{osc}}{\gamma}$
$\boldsymbol{\omega}_{eff}$
$\mathbf{B}_{eff}$
$\boldsymbol{\mu}$
$\mathbf{B}_1$
x
$-y$

Figure B. Effective Magnetic Field in the Rotating Frame
The motion of the moment about $\mathbf{B}_{eff}$ is easier to describe in the frame rotating about the *z*-axis at the same frequency ω_{osc} as the oscillating field $\mathbf{B}_1$, since then $\mathbf{B}_1$ is static. $\mathbf{B}_0$ is reduced by the amount ω_{osc}/γ.

where ω_L is called the Larmor frequency and is measured in radians per second. Equation (4) is the single most important equation of magnetic resonance. It says that the frequency of precession about a magnetic-field line is proportional to both the magnitude of the magnetic field and the gyromagnetic ratio. Interestingly, as derived in the equations accompanying the figure, the frequency is independent of the angle θ that specifies the orientation of the magnetic moment. The Larmor frequency enables us to identify the particle because the gyromagnetic ratio is distinct for electrons and different nuclei. The Larmor frequency ($\omega_0/2\pi$) for an electron is about 28 gigahertz per tesla (MHz/T) and for a proton, roughly 45 MHz/T.

The moments precessing in the applied field can be manipulated in several ways. One common method is pulsed magnetic resonance. For a short period, we apply an oscillating magnetic field along the x-axis, $\mathbf{B}_1 = B_1 \cos(\omega_{osc} t)\ \hat{x}$, where $B_1 << B_0$. The moment will begin to precess around a time-dependent total magnetic field consisting of $\mathbf{B}_0$ plus $\mathbf{B}_1$. This complicated motion can be better understood by examining the moment in a reference frame that rotates around the z-axis with frequency $\omega_{osc,}$ the same frequency as $\mathbf{B}_1$. In the rotating frame, $\mathbf{B}_1$ becomes a static field and the precession frequency about the z-axis is reduced: $\omega_L \to \omega_L - \omega_{osc}$.

Phenomenologically speaking, in the rotating frame the magnetic moment "sees" an effective field of magnitude

$$\begin{aligned}\mathbf{B}_{\text{eff}} &= \left(B_0 - \frac{\omega_{\text{osc}}}{\gamma}\right)\hat{z} + B_1\hat{x} \\ &= \frac{1}{\gamma}\left(\omega_{\text{L}} - \omega_{\text{osc}}\right)\hat{z} + B_1\hat{x} \quad , \qquad (5)\end{aligned}$$

which is illustrated in Figure B. Equation (5) tells us that, when the frequency of the $\mathbf{B}_1$ equals the Larmor frequency, namely, at the resonance condition $\omega_{\text{L}} = \omega_{\text{osc}}$, the effective field has no z-component. Only the $\mathbf{B}_1$ field remains, and the moment will precess around the x-axis at an angular frequency ω_1 set by the magnitude of $\mathbf{B}_1$, namely, $\omega_1 = \gamma B_1$. Thus in the laboratory, we can rotate a moment about the x-axis by setting the frequency of $\mathbf{B}_1$ to the Larmor frequency. We control the rate of rotation by adjusting the field strength and the amount of rotation by restricting the length of time that the $\mathbf{B}_1$ field is applied.

Pulsed magnetic resonance can be used to manipulate a qubit. Suppose a qubit state is defined by the nuclear spin orientation such that the spin aligned parallel to $\mathbf{B}_0$ represents the state $|0\rangle$ whereas the spin aligned antiparallel to the field represents the state $|1\rangle$. We send a current pulse through an inductive coil to create the field B_1. If the pulse is timed to last for one-half of a precession period, or $t = \pi/\omega_1$, then the spins will rotate around the x-axis for π radians, or by 180°. If the qubit was initially in the $|0\rangle$ state, it would now be in the $|1\rangle$ state. Similarly, we can pulse the current for a time $t = \pi/(2\omega_1)$—a so-called $\pi/2$ pulse—and rotate the qubit into an equal superposition of the $|0\rangle$ and $|1\rangle$ states, namely the state $1/\sqrt{2}\,(|0\rangle + |1\rangle)$. (See Figure C.)

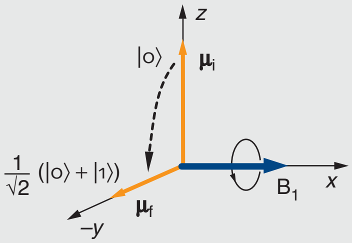


Figure C. Pulsed Magnetic Resonance

When ω_{osc} is made equal to ω_{L}, a moment will begin to rotate about the *x*-axis. We place a qubit into an equal superposition of logical states by rotating the moment through 90° with a $\pi/2$ pulse, in which B_1 is turned on for a time $t = \pi/(2\omega_1)$, where $\omega_1 = \gamma B_1$.

We can also make moments rotate continuously about the x-axis. In a process known as cyclic adiabatic inversion, we sweep ω_{osc} through a range that includes the Larmor frequency. When we start the sweep, $\omega_{\text{osc}} << \omega_{\text{L}}$. According to Equations (5), there is little cancellation of the static field $\mathbf{B}_0$, and $\mathbf{B}_{\text{eff}}$ will lie substantially along the z-axis. As the frequency approaches ω_{L}, there is more cancellation, and $\mathbf{B}_{\text{eff}}$ begins to rotate toward the x-axis. When $\omega_{\text{osc}} = \omega_{\text{L}}$, $\mathbf{B}_{\text{eff}}$ points along the x-axis. Continuing to sweep the frequency to $\omega_{\text{osc}} >> \omega_{\text{L}}$ will eventually cause $\mathbf{B}_{\text{eff}}$ to point along the $(-z)$-axis. If ω_{osc} is swept slowly enough (the adiabatic condition), the moments will continue to precess around $\mathbf{B}_{\text{eff}}$ and will follow its rotation in the x-z plane from $+z$ to $-z$ (See Figure D). Reversing the sweep will cause $\mathbf{B}_{\text{eff}}$ to rotate backwards. By continuously sweeping ω_{osc} back and forth through the resonance frequency, we effectively make the spins rotate continuously around the y-axis.

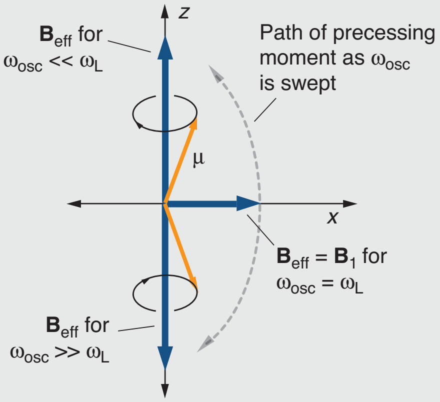

z
$\mathbf{B}_{\text{eff}}$ for $\omega_{\text{osc}} << \omega_{\text{L}}$
Path of precessing moment as ω_{osc} is swept
μ
x
$\mathbf{B}_{\text{eff}} = \mathbf{B}_1$ for $\omega_{\text{osc}} = \omega_{\text{L}}$
$\mathbf{B}_{\text{eff}}$ for $\omega_{\text{osc}} >> \omega_{\text{L}}$

Figure D. Cyclic Adiabatic Inversion

$\mathbf{B}_{\text{eff}}$ rotates about the *y*-axis when ω_{osc} is swept through the resonance frequency ω_{L}. If ω_{osc} changes slowly, the moment continues to precess about $\mathbf{B}_{\text{eff}}$ and we can rotate the moment about the *y*-axis.

Cyclic adiabatic inversion provides one of the mechanisms by which we detect electron moments with a magnetic resonance force microscope (MRFM). A small number of moments are in resonance with $\mathbf{B}_0$, $\mathbf{B}_1$, and the gradient field produced by the magnetic tip at the end of the MRFM cantilever. We use cyclic adiabatic inversion to selectively rotate those moments, thus producing a tiny oscillating magnetization within the sample that in turn produces an oscillating force on the MRFM cantilever. By adjusting the rate at which we sweep ω_{osc}, we can match the forcing frequency to the cantilever's resonant frequency, and even a small number of moments can drive the cantilever into a detectable oscillation.

Continued from page 287

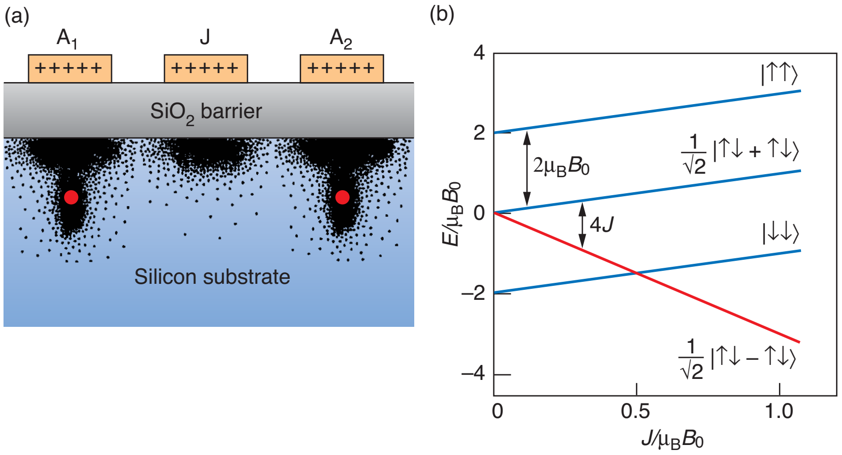

(a)
A_1
J
A_2
+++++
+++++
+++++
SiO_2 barrier
Silicon substrate
(b)
4
2
0
−2
−4
$E/\mu_B B_0$
$|\uparrow\uparrow\rangle$
$\frac{1}{\sqrt{2}}|\uparrow\downarrow + \uparrow\downarrow\rangle$
$2\mu_B B_0$
$4J$
$|\downarrow\downarrow\rangle$
$\frac{1}{\sqrt{2}}|\uparrow\downarrow - \uparrow\downarrow\rangle$
0
0.5
1.0
$J/\mu_B B_0$

Figure 3. Coupling Two Qubits with a J-Gate

(a) When the gates A_1 and A_2 are appropriately biased, application of a small positive voltage to the J-gate lowers the potential barrier between adjacent donor sites and turns on an electron-spin exchange interaction between qubits, as seen in this cartoon. The electrons interact with the nuclei through the hyperfine interaction; hence, the two nuclear spins become coupled to each other. (b) The graph shows energy levels of the coupled electron-spin system as a function of the electron-spin exchange coefficient *J*, which can be modified by voltage applied to the J-gate. The electrons couple their spins to form three states with *S* = 1 (shown in blue) and one with *S* = 0 (shown in red). For $J/\mu_B B_0 < 0.5$, the electrons occupy the lowest energy *S* = 1 state $|\downarrow\downarrow\rangle$. Two-qubit operations are performed in this low *J* regime.

interaction. The exchange coupling coefficient J is proportional to the overlap between the wave functions of the two donor electrons, so its strength is a function of the J-gate voltage.

We first examine the coupled electron-spin states by ignoring (for a moment) the contribution of the nuclear spins to H^1 and H^2 in Equation (4). The effect of the spin exchange interaction is to create coupled electron-spin states, three with total spin $S = 1$ and one with total spin $S = 0$. The respective wave functions are

$$S = 1 \quad |\uparrow\uparrow\rangle \;, \quad 1/\sqrt{2}\,|\uparrow\downarrow + \downarrow\uparrow\rangle \;, \text{ and} \quad |\downarrow\downarrow\rangle \;,$$

$$S = 0 \quad |S\rangle = 1/\sqrt{2}\,|\uparrow\downarrow - \downarrow\uparrow\rangle \;.$$

In the absence of a magnetic field, the energy difference between the states with $S = 1$ and $S = 0$ is $4J$, an amount known as the exchange energy. In the presence of the magnetic field $\mathbf{B}_0$ that permeates the quantum computer, the $|\uparrow\uparrow\rangle$ and $|\downarrow\downarrow\rangle$ states are Zeeman-split around zero by an amount $\pm 2\mu_B B_0$, and the energies of the coupled electron-spin states vary as a function of J, as seen in Figure 3(b). Notice that the lowest-energy $S = 1$ state and the $S = 0$ state cross when the exchange energy becomes equal to the Zeeman splitting, that is, when $4J = 2\mu_B B_0$. We exploit that crossing in a qubit readout scheme discussed later.

We now consider the nuclear spin states. Conceptually, for every coupled electron-spin state, there are four possible orientations of the two nuclear spins, corresponding to the uncoupled ($J = 0$) nuclear states $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. Thus, there are sixteen nuclear spin states in all. Formally, we must use Equation (4) to find the energies and eigenfunctions of all sixteen.[3] If we focus only on those states associated with the electron ground state $|\downarrow\downarrow\rangle$ and assume $4J < 2\mu_B B_0$, then in order of decreasing energy, the coupled nuclear-spin states are the following:

$$|11\rangle \;,\quad |\Phi_+\rangle = 1/\sqrt{2}\,|10 + 01\rangle \;,\quad |\Phi_-\rangle = 1/\sqrt{2}\,|10 - 01\rangle \;, \text{ and} \quad |00\rangle \;. \tag{5}$$

The electron-spin exchange interaction shifts the energy of the $|\Phi_-\rangle$ state below that of $|\Phi_+\rangle$ by an amount

$$\Delta E_J = \hbar\omega_J = 2A^2\left(\frac{1}{\mu_B B_0 - 2J} - \frac{1}{\mu_B B_0}\right) \;, \tag{6}$$

where ω_J is the nuclear exchange frequency. For $B_0 = 2T$ and $4J = 0.124$ milli-electron-volt (meV), ω_J has a value of about (2π)75 kilohertz, a frequency that approximates the rate at which binary operations can be performed on the computer.

The spin exchange interaction causes the wave functions of Equation (5) to evolve and rotate between spin states. One possible result is that the nuclear spins undergo a coordinated swapping of states: $|q_1 q_2\rangle \rightarrow |q_2 q_1\rangle$ (see the box "The Swap" on the facing page). Thus, the spin exchange interaction should automatically implement the logical two-qubit **swap** gate.

Of more interest is the **cnot** gate, which along with single-qubit operations, forms a universal set of gates from which any quantum algorithm can be executed. In the Kane system, the **cnot** corresponds to the conditional rotation of a target spin by 180°, provided the control spin is in the state $|1\rangle$. In principle, it can be realized by subjecting the wafer to a transverse magnetic field $\mathbf{B}_1$ and applying voltages to the A- and J-gates (Goan and Milburn 2000).

---

[3] The energy differences between the four nuclear states associated with each electron state are very small. A graph of the nuclear-electron energy levels would look identical to Figure 3(b), except that under high magnification one would see that each line consists of four closely spaced lines.

Suppose that the two electron spins are initially uncoupled ($J = 0$) and that the hyperfine coupling constants A_1 and A_2 of the two donor atoms are equal ($A_1 = A_2$). In that case, biasing the A-gates such that $A_1 > A_2$ distinguishes the control qubit from the target. We then turn on the spin exchange interaction ($J > 0$) and slowly make A_1 equal to A_2. The result would be that the uncoupled qubit state $|10\rangle$ evolves adiabatically into the state $|\Phi_+\rangle = 1/\sqrt{2}\,|10 + 01\rangle$, and $|01\rangle$ evolves into $|\Phi_-\rangle = 1/\sqrt{2}\,|10 - 01\rangle$. When A_1 equals A_2, the energy splitting between the two states is given by Equation (6). The states $|11\rangle$ and $|00\rangle$ are unaffected by the procedure.

We next apply a linearly polarized oscillating field B_1 with frequency that is resonant with the $|11\rangle - |\Phi_+\rangle$ energy difference. The field is left on until the $|11\rangle$ state has rotated into the $|\Phi_+\rangle$ state and vice versa. By executing a reverse of the sequence of steps performed at the beginning of the operation, we adiabatically transform the $|\Phi_+\rangle$ and $|\Phi_-\rangle$ states back into $|10\rangle$ and $|01\rangle$, respectively. We effect the change

$$\begin{aligned} |00\rangle &\to |00\rangle\ , \\ |01\rangle &\to |01\rangle\ , \\ |10\rangle &\to |11\rangle\ , \text{ and} \\ |11\rangle &\to |10\rangle\ . \end{aligned}$$

That is, the only qubits that are flipped are those in which the control qubit is in the state $|1\rangle$. Thus, we expect to be able to perform the **cnot** operation.

The Swap

Before the J-gate is turned on, the two nuclear spins are uncoupled, and each is described by the following energy eigenstates: $|\Psi_1\rangle = |00\rangle$, $|\Psi_2\rangle = |01\rangle$, $|\Psi_3\rangle = |10\rangle$, and $|\Psi_4\rangle = |11\rangle$. Once the J-gate is turned on, the coupled eigenstates are $|00\rangle$, $|\Phi_-\rangle = 1/\sqrt{2}\,|10 - 01\rangle$, $|\Phi_+\rangle = 1/\sqrt{2}\,|10 + 01\rangle$, and $|11\rangle$.

Suppose the uncoupled nuclear spins were originally in the state $|\Psi_2\rangle = |01\rangle$, and then voltage was applied quickly to the J-gate. In terms of the eigenstates of the coupled system, the system finds itself in the state

$$|\Psi_2\rangle = 1/\sqrt{2}\,(|\Phi_+\rangle - |\Phi_-\rangle)\ . \tag{1}$$

The time evolution of this wave function (up to an overall phase) is given by

$$|\Psi_2(t)\rangle = 1/\sqrt{2}\,(|\Phi_+\rangle - e^{-i\omega_J t}\,|\Phi_-\rangle)\ , \tag{2}$$

where ω_J is the nuclear exchange frequency. After a time $t = \pi/\omega_J$, the wave function will evolve into

$$|\Psi_2(\pi/\omega_J)\rangle = 1/\sqrt{2}\,(|\Phi_+\rangle + |\Phi_-\rangle) = |\Psi_3\rangle\ . \tag{3}$$

That is, the system will have evolved from the state $|01\rangle$ to the state $|10\rangle$. The spins will have swapped. If we quickly remove the voltage from the J-gate, the two-spin system will remain in the state $|10\rangle$.

Approaches to Readout

One can evaluate the result of a quantum computation only by reading the final state, $|0\rangle$ or $|1\rangle$, of a qubit. Likewise, the ability to determine the state of a given qubit is critical to initializing the quantum register. Ideally, we would read the qubit state directly by measuring the donor atom's nuclear-spin state. But direct detection of a single nuclear spin is currently impossible and may forever be out of our grasp. (The strength of the coupling between a magnetic field and the nuclear spin is set by the magnitude of the nuclear magneton μ_n, which is very small.) We are therefore forced to find some other means of reading out the qubit state.

The potential solution is to correlate the nuclear spin states of a target atom with the electron spin and to find some way of determining the electron spin state. We are currently pursuing two distinct detection schemes, one involving a single electron transistor (SET) and the other, a magnetic resonance force microscope (MRFM). Both approaches require that we push the respective technologies beyond the current state of the art.

Single-Charge Measurement Using SETs. The idea behind this technique, first described by Kane (1998), is to couple the target qubit q_t to a readout qubit q_r by a J-gate, and then infer the state of q_t by monitoring the donor electrons of the coupled system. If q_t is in the state $|0\rangle$, we can cause both electrons to become localized around the readout atom (they would occupy the so-called D^- state). If q_t is in the state $|1\rangle$, each donor electron would remain bound to its respective atom. An SET would be used as an ultrasensitive electrometer to determine whether one or two electrons were bound to the readout atom.

The procedure can be understood with reference to Figure 4(a), which shows the coupled nuclear-spin states in the vicinity of the electron spin crossing. As discussed in the previous section, for $J/\mu_B B_0 < 0.5$, the lowest-energy electron spin state is the S = 1 state $|T\rangle = |\downarrow\downarrow\rangle$, but for $J/\mu_B B_0 > 0.5$, the $S = 0$ state $|S\rangle = 1/\sqrt{2}\ |\uparrow\downarrow - \downarrow\uparrow\rangle$ assumes the lower energy.

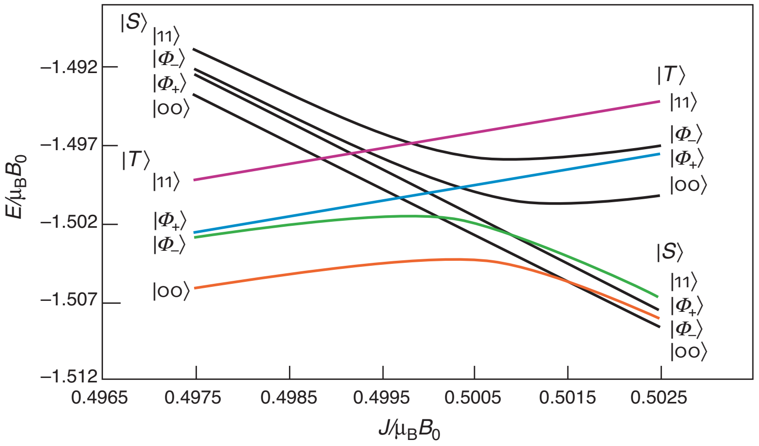
Figure 4(a) shows what happens to the eight lowest-energy nuclear-spin states as the electron-spin states cross. Focusing on the four states initially associated with $|T\rangle$, we see that after the crossing, the two higher-energy nuclear states $|11\rangle$ and $|\Phi_+\rangle$ remain coupled to $|T\rangle$, while the two lower-energy states $|00\rangle$ and $|\Phi_-\rangle$ get coupled to $|S\rangle$. In other words, as we increase J, we can adiabatically evolve both the nuclear- and electron-spin systems. If the target qubit was originally in the state $|0\rangle$, then regardless of the state of the readout qubit, the electrons will evolve into the $S = 0$ spin state. If q_t is originally in the state $|1\rangle$, the electrons will remain in the lowest energy $S = 1$ spin state. The sequence of steps, similar to those used to implement the **cnot** gate, is outlined in Figure 4(b).

We next use the fact that the only two-electron bound state of a phosphorous atom in silicon is the D^- state with total spin $S = 0$. As seen in Figure 4(c), we would bias the A- and J-gates to create an electric field between the two donor atoms. If the electrons are in the $S = 0$ state, the target electron can transfer to the readout atom, and we would know that the target atom was initially in the state $|0\rangle$.

An SET would be used to detect the presence of the second donor electron about the readout atom. In many ways, an SET is like an ordinary transistor, in that a gate electrode moderates the current flowing between a source and drain electrodes. The difference is that between the SET's source and drain lies an extremely small metallic island, which is isolated from each electrode by small patches of insulating material. The insulator acts as a tunnel junction. For current to flow, electrons must tunnel from the source to the island and then from the island to the drain. The tunneling current is greatly affected by the

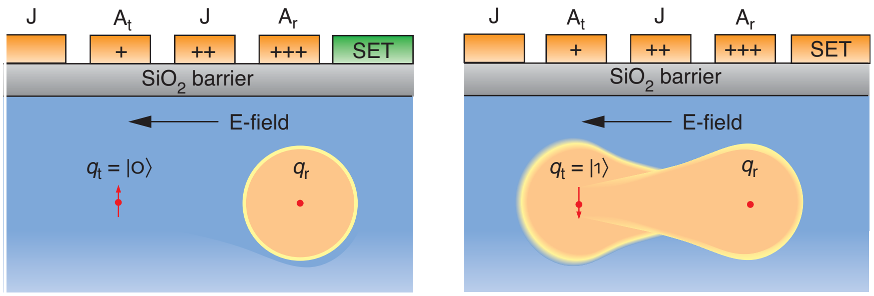
(a)

	Electron States	Nuclear States
$S = 1$	$\lvert T\rangle = \lvert\downarrow\downarrow\rangle$	$\lvert\Phi_+\rangle = 1/\sqrt{2}\ \lvert 10 + 01\rangle$
$S = 0$	$\lvert S\rangle = 1/\sqrt{2}\ \lvert\uparrow\downarrow - \downarrow\uparrow\rangle$	$\lvert\Phi_-\rangle = 1/\sqrt{2}\ \lvert 10 - 01\rangle$



(b) States Adiabatically Evolved through the Crossing

	$J/\mu_B B_0 < 0.5$, $A_t - A_r >> \hbar\omega_J$		$J/\mu_B B_0 < 0.5$, $A_t = A_r$		$J/\mu_B B_0 > 0.5$, $A_t = A_r$
q_t initially in state $\lvert 1\rangle$	$\lvert T\rangle\lvert 11\rangle$	→	$\lvert T\rangle\lvert 11\rangle$	→	$\lvert T\rangle\lvert 11\rangle$
	$\lvert T\rangle\lvert 10\rangle$	→	$\lvert T\rangle\lvert\Phi_+\rangle$	→	$\lvert T\rangle\lvert\Phi_+\rangle$
q_t initially in state $\lvert 0\rangle$	$\lvert T\rangle\lvert 01\rangle$	→	$\lvert T\rangle\lvert\Phi_-\rangle$	→	$\lvert S\rangle\lvert 11\rangle$
	$\lvert T\rangle\lvert 00\rangle$	→	$\lvert T\rangle\lvert 00\rangle$	→	$\lvert S\rangle\lvert\Phi_-\rangle$



(c) Tunneling ($q_t = |0\rangle$)

(d) No Tunneling ($q_t = |1\rangle$)

Figure 4. Single-Electron Transistor (SET) Readout Scheme
(a) The graph shows the eight lowest-energy nuclear-spin states for the coupled target and readout qubits $|q_t, q_r\rangle$ in the region where the $S = 0$ and the lowest energy $S = 1$ electron-spin states cross. (b) We can adiabatically evolve the nuclear-electron states by biasing the J- and A-gates, as seen in this (partial) sequence of steps. The electrons are initially in the $S = 1$ state $|T\rangle$. If q_t was initially in the $|1\rangle$ state, then the electrons will remain in $|T\rangle$ regardless of the state of q_r. If initially $q_t = |0\rangle$, then at the end of the sequence, the electrons will be in the $S = 0$ state $|S\rangle$. (c) Only the two electrons in the $|S\rangle$ state can bind to a single phosphorous atom in silicon. Given a suitable biasing of the gate electrodes, we can try to induce an electron to tunnel to a readout qubit q_r. If the tunneling is successful, the electrons were in the $|S\rangle$ state, and $q_t = |0\rangle$. The tunneling current would be detected by an SET located near q_r. (d) If no tunneling occurs, the two electrons were in the $|T\rangle$ state, and hence $q_t = |1\rangle$.

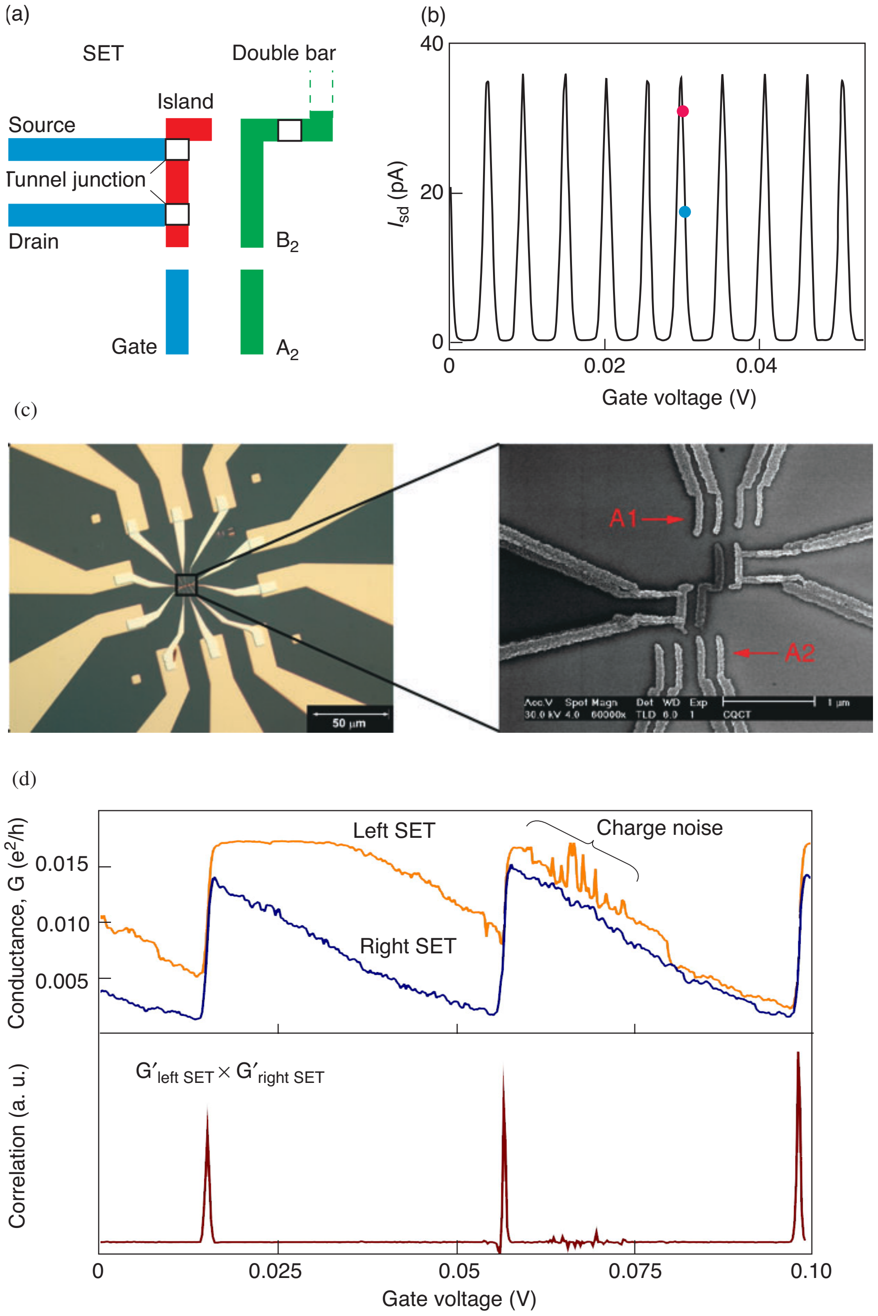


Figure 5. Twin SET Device for Readout Simulation

(a) The figure shows one half of a twin SET/ double-bar test device. The SET consists of a small metal island connected to source and drain leads by tunnel junctions and a gate electrode that is capacitively coupled to the island. Electrons can pass from source to drain only by tunneling through both junctions. The SET is located next to a metal bar B_2, which is isolated from the other bar B_1 by a tunnel junction.
(b) The tunneling current I_{sd} in the SET is strongly influenced by a change in the local charge distribution. If the gate voltage is originally biased at V_0, (blue dot), then a change in the local charge distribution effectively modifies it to $V_0 - \delta$, and the source-drain current will change dramatically (red dot). (c) This is an image of the twin-SET test device obtained with a scanning electron microscope. The image to the right is a magnified version of the central region. The twin-SET device is fabricated by a double-angle evaporation process, which replicates each of the features. Unequal voltage on A_1 and A_2 causes an electron to tunnel from one bar to the next. (d) The movement of charge is detected as a change in the source/drain conductance in both SETs simultaneously. The two signals can be correlated to discriminate the charge transfer signal from reproducible charge noise or from random noise events.

capacitive coupling between the gate and the island. This means that for particular voltage biases on the gate, source, and drain, current flow through the SET becomes exquisitely sensitive to minute changes in the charge distribution of the local environment. The presence of a single additional electron is readily detectable as a change in the SET's source/drain conductance.

We have developed several readout simulation devices to test the properties of our SETs built in house. In the device seen in Figure 5, two thin metal bars, isolated from each other by a tunnel junction, substitute for the phosphorous atoms. Control gates are used to electrostatically "push" single electrons from one bar to the next. The two SETs are then used to detect the change in the charge distribution due to the discrete, single-electron tunneling events. Those events cause the output of both SETs to change abruptly. In contrast, signals due to unwanted charge noise (reproducible fluctuations in the conductance versus voltage curve) tend not to affect both SETs simultaneously. By correlating the outputs of the two SETs, we are able to clearly identify the single-charge transfer events and reject spurious signals that would interfere with the readout.

Other factors, however, also need to be considered before we use an SET in a qubit readout scheme. Suppose the target qubit q_t is initially in the $|1\rangle$ state. Then, for high values of J, the coupled electrons will remain in the higher-energy state $|T\rangle$ (refer to Figure 4). This means that the coupled atomic system could lower its energy if one of the polarized electrons "relaxed" and flipped to form the state $|S\rangle$. The electron would then transfer to the readout qubit, and we would erroneously deduce that q_t was initially in the $|0\rangle$ state! Recent results suggest that the spin relaxation time is of the order of milliseconds. We must therefore pull information out of the SET on an even shorter time scale. We must be able to determine that a change occurred in the SET conductance at a time t_0, rather than a few milliseconds after t_0.

Unfortunately, that is difficult to do with an ordinary SET. The measurements are made at liquid helium temperatures, and the SET, sitting in a cryostat, must somehow be connected to the outside world. The capacitance of the connecting cables is fairly large, and when combined with the intrinsic resistance of the SET, produces a resistance-capacitance (or R-C) time constant for the device that is longer than the spin relaxation time. Information about the SET conductance takes too long to propagate to the outside world.

The solution to this problem is to develop a fast readout SET (Schoelkopf et al. 1998). Known as an rf SET, it is basically an ordinary SET coupled to an impedance-matching circuit that negates the effects of the external capacitance. We have recently developed a very sensitive reflection-mode rf SET that operates at 430 megahertz. It can detect the movement of a single electron in the device shown in Figure 5 in about 1 microsecond. For a system containing discrete phosphorous atoms, the readout time would likely increase to about 100 microseconds, but that is still sufficient for the readout approach discussed in this section.

The MRFM. The second approach to readout is to perform a direct measurement of the spin state of the electron surrounding the qubit and thereby infer the qubit state. To do so, we are developing an MRFM, which combines the versatility and chemical specificity of magnetic resonance with the high-resolution and ultrahigh sensitivity of an atomic force microscope (AFM). The key feature of the MRFM is that only spins contained within a defined area in the sample—the so-called sensitive slice—contribute to the detected signal. Because the location and size of that slice can be controlled, there is selective sensitivity to deeply buried structures.

Our MRFM, developed at Los Alamos in collaboration with Michael Roukes of Caltech, is illustrated in Figure 6. The microscopic, sharp-pointed magnetic tip is bonded to the end of a tiny cantilever. As in an ordinary AFM, the tip is scanned over a sample, and signals are recorded at every point. In our instrument, however, the magnetic field from the tip $\mathbf{B(r)}$ interacts with all the electron spins in the substrate through the magnetic gradient force, $\mathbf{F(r)} = (\mathbf{m}\cdot\nabla)\mathbf{B(r)}$, where $\mathbf{m}$ is the net magnetization of the spins. Depending on the spin orientation, the force on the tip is either repulsive or attractive. The net orientation of the electron spins in the sample, therefore, causes a tiny deflection of the cantilever.

We interact with only a subset of the spins through magnetic resonance. The sample is immersed in a static magnetic field $\mathbf{B}_0 = B_0\hat{z}$, so the precession frequency of the spins around the magnetic-field lines is proportional to $\mathbf{B}_0$ plus the z-component of $\mathbf{B(r)}$, that is, the total magnetic field in the z-direction**.** Magnetic resonance comes into play when we subject the spins to an oscillating magnetic field $\mathbf{B}_1$ that is aligned in the x-direction. Because the magnitude of $\mathbf{B(r)}$ decreases rapidly with distance, only those spins that are at the correct distance from the tip are in resonance with the $\mathbf{B}_1$ field. Spins that are too close to the tip will have a higher resonant frequency; those that are farther away, a lower frequency. Thus, for a given field gradient and a fixed cantilever position, a resonance frequency becomes correlated with positions inside the sample. With reference to Figure 6, all spins that lie within a small, hemispherical shell beneath the tip (the sensitive slice) have the same resonance frequency.

To detect those spins, we use the technique of cyclic adiabatic inversion, discussed in the box "Spin Manipulation with Magnetic Resonance" on page 288. In essence, we continuously rotate the selected spins at the resonant frequency of the cantilever. The continuous up and down reorientation of the spins creates an oscillating force on the tip that amplifies the cantilever's natural up-down motion. The situation is analogous to pushing a child's swing at its natural frequency of oscillation: with each push, the amplitude of the motion becomes larger. After thousands of spin rotations, the amplitude of the cantilever's up-down motion has increased by about an angstrom, which is large enough to be detected with a fiberoptic interferometer. The fiber sits slightly above the back of the cantilever, and laser light sent down the fiber interferes with itself as it reflects from both the cantilever and the fiber's end. By monitoring changes in the interference pattern, we can detect the oscillations.

The orientation of the nuclear spins can be inferred from the frequency at which the electron spin resonance

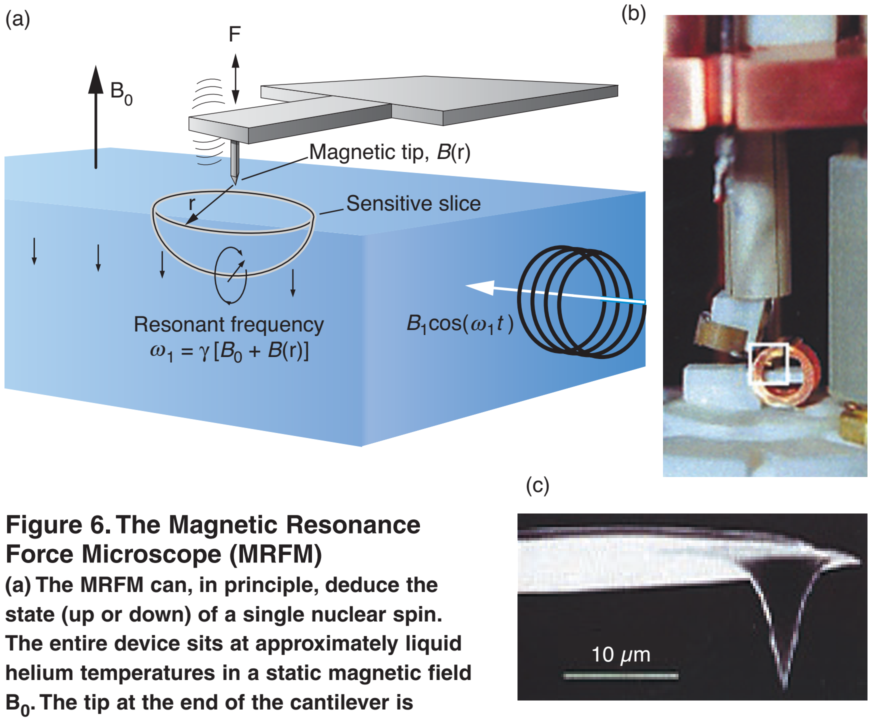


Figure 6. The Magnetic Resonance Force Microscope (MRFM)
(a) The MRFM can, in principle, deduce the state (up or down) of a single nuclear spin. The entire device sits at approximately liquid helium temperatures in a static magnetic field B_0. The tip at the end of the cantilever is coated with a magnetic material that generates a magnetic field B(r) that changes rapidly with the distance r. The interaction between the electron spins in the sample and the magnetic field gradient due to B(r) produces a force that deflects the cantilever. We interact with only a small subset of spins, located within a hemispherical shell of radius r_1, by subjecting the sample to an oscillating magnetic field $B_1\cos(\omega_1 t)$, where $\omega_1 = \gamma[B_0+B(r_1)]$. By using the technique of cyclic adiabatic inversion, we can cause the spins to oscillate between the up and down states at the cantilever resonance frequency, thus driving the cantilever into measurable oscillation. We detect the oscillation with an optical device. The electron-resonance frequency can then be correlated with a nuclear spin orientation. (b) The MRFM tip assembly and sample mount are shown in this photo. The vertical tube is a piezo scanning tube, which moves the tip over the sample, while the circular feature is the induction coil that produces B_1. The white box highlights the magnetically coated tip, shown under high magnification in (c).

occurs. Because of the hyperfine interaction, the resonance frequency of an electron spin flip depends on the nuclear-spin state. Considering the hyperfine states of a single qubit, the $|0\downarrow\rangle$ to $|0\uparrow\rangle$ transition has a different energy than the $|1\downarrow\rangle$ to $|1\uparrow\rangle$ transition, and thus there are two resonance frequencies for an electron spin transition. Measurement of, say, the higher resonance frequency would correspond to the nuclei in the sample being aligned with the $\mathbf{B}_0$ field.

The discussion so far has centered on detecting many nuclear spins, but to read out the result of a quantum computation, we need to measure a single nuclear spin. That such measurement is at all possible is due to the exceedingly high spatial resolution of the MRFM, which is determined by the thickness Δz of the hemispherical shell. The thickness is inversely proportional to the magnitude of the field gradient:

$$\Delta z \cong \frac{\Delta\omega_r}{\gamma\left|\nabla_z \mathbf{B}(\mathbf{r})\right|} \quad , \qquad (7)$$

where γ is the gyromagnetic ratio and $\Delta\omega_\mathrm{r}$ is the linewidth of the resonant electron-spin transition that is being driven by the MRFM. For phosphorus atoms in silicon, $\Delta\omega_\mathrm{r}/\gamma$ is on the order of 1 milligauss. A field gradient of about 10^5 tesla per meter (T/m) will then produce a thickness that is much less than 1 angstrom, even when the hemispherical shell extends several hundred angstroms beneath the substrate surface. In that case, the sensitive slice would be so thin that only a single donor electron would be in resonance with the MRFM probe.

We have conducted numerous experiments to measure the field gradient of our specialized magnetic tips. With the tip about 2 micrometers from the surface, we have measured a field gradient approaching 10^4 T/m (see Figure 7). From this value, we estimated Δz and the volume of our hemispherical shell. Then, knowing the spin density of the sample, we estimated the number of spins that contribute to the signal. For the data shown in Figure 7, the number is between one thousand and ten thousand electron spins.

Because the field gradient increases nearer to the tip, sensitivity should be greater if the tip is closer to the surface. But mechanical and thermal noise also deflect the tip and cantilever. As we begin to interact with fewer spins, the "signal" force due to spins eventually becomes less than the "noise" force due to unwanted sources. By equating expressions for the signal force to the noise force, we can derive an expression for the minimum detectable magnetic moment, m_min, needed to give a signal to noise of 1:

$$m_\mathrm{min} = \frac{1}{\left|\nabla_z \mathbf{B}(\mathbf{r})\right|}\sqrt{\frac{2kk_\mathrm{B}T\Delta\nu}{\pi Q f_\mathrm{c}}} \quad . \qquad (8)$$

In Equation (8), k_B is the Boltzmann constant; T, the temperature; and $\Delta\nu$, the detection bandwidth. The other parameters describe the cantilever: its force constant k, resonant frequency f, and quality factor Q. The three key

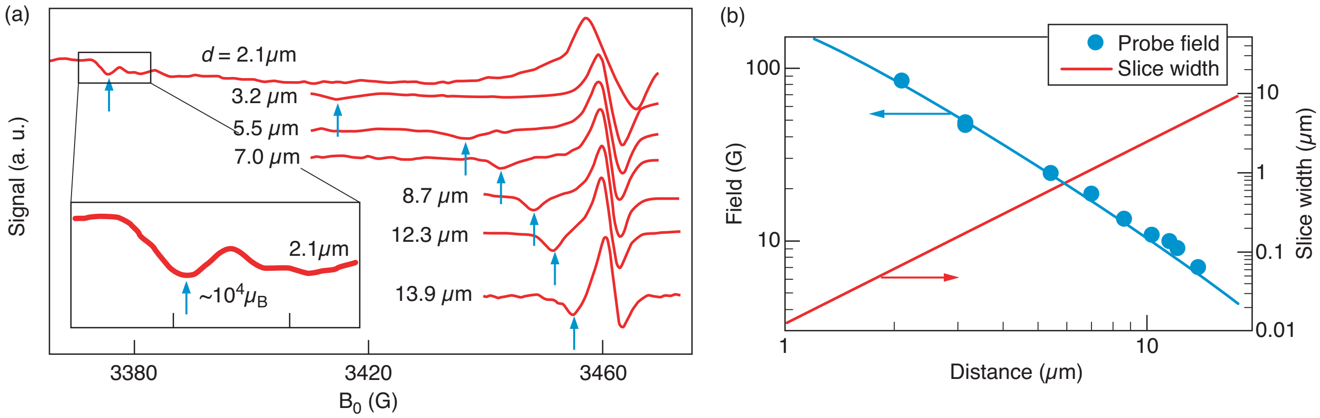


Figure 7. Sensitivity of the MRFM

(a) If the magnetic tip is kept at a fixed distance *d* from the sample, then lowering the value of the magnetic field corresponds to sweeping the sensitive slice upwards, toward the surface. At some point, the slice leaves the sample, and the resonance condition changes dramatically. That change is seen in the derivative of the MRFM signal as a dip, indicated by the arrows. (b) By following the dip as a function of tip height, we can measure the tip's magnetic field. From the field gradient, we then calculate the width of the sensitive slice using Equation (7) in the text. Knowing the spin density within the sample, we use the slice width to deduce how many spins produced the signal and thereby infer the MRFM sensitivity.

parameters that we can optimize are the field gradient, temperature, and quality factor.

We believe that the sensitivity of the MRFM is currently limited by surface contamination on the sample. As the tip approaches the surface, the contamination acts like a viscous force that damps the oscillatory motion—that is, it lowers the Q in Equation (8). To solve this problem, we are upgrading the equipment so that the sample be transferred from a surface preparation chamber into the microscope without leaving the ultrahigh vacuum environment. The system will also be cooled to temperatures between 250 and 300 millikelvins in a helium-3 dilution refrigerator, a technique that is compatible with maintaining the sample under ultrahigh vacuum.

Detection of a single electron moment requires that

$$m_{\min} = 1\ \mu_B \cong 10^{-23}\ \text{joule/tesla} \quad . (9)$$

Given a field gradient of 10^5 T/m, the signal force on the cantilever is approximately 10^{-18} newton (the weight of approximately two million phosphorus atoms). We believe that an upgraded, low-temperature microscope will allow us to observe the magnetic resonance signal of a single electron spin.

SSQC Fabrication Progress

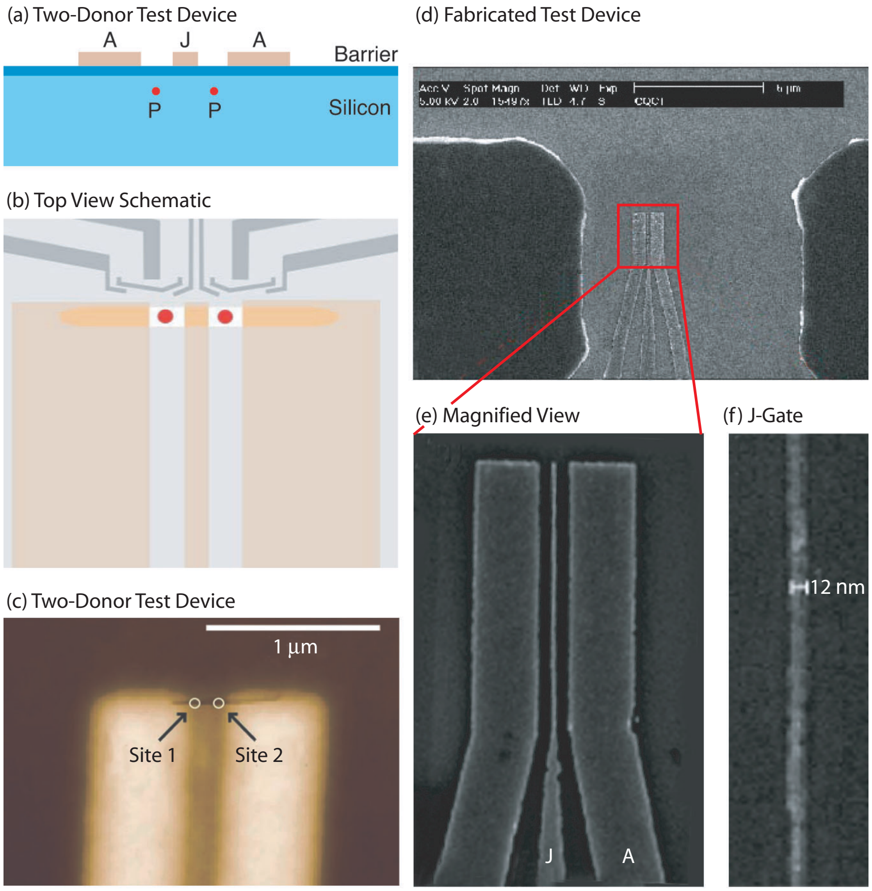
Implementing our quantum-computing scheme requires that we produce a very regular array of phosphorus atoms in pure silicon, in which each donor is located precisely beneath a metal A-gate on the surface. The spacing between adjacent phosphorus donors is chosen to ensure that the electron-spin exchange interaction is minimal when there is no voltage on the J-gate lying between the donors. We want the two electron wave functions to overlap, but only slightly. Calculations (Goan and Milburn 2000) indicate that a separation of 10 to 20 nanometers between donors is required.

A nominal donor spacing of 20 nanometers translates into gate structures that are less than 10 nanometers in width. Fabricating a highly regular metal array on that scale, even with state-of-the-art techniques, is at the limits of the electron-beam techniques used in making conventional electronics. That problem, however, pales when compared with the difficulties we face in making a precisely aligned array of phosphorous donors that is buried under layers of silicon. The difficulties have led us to pursue two different fabrication strategies, known as the top-down and bottom-up approaches.

In the top-down approach, phosphorous atoms will be implanted by ion bombardment into specific sites on the silicon wafer. Because the ion scatters as it slows down in the silicon, we will not know the exact location of the donors, only that they will lie within close range of the defined implantation area. The top-down approach provides a rapid means to demonstrate proof of principle and allows us to fabricate a two-donor device that can be used to test readout strategies and, possibly, quantum operations. Scaling this approach to large numbers of qubits will be challenging, because of the irregular spacing of the donor array.

In contrast, the bottom-up approach will use a scanning tunneling microscope (STM) with which to place phosphorous atoms on a clean silicon surface in a precisely arranged array. The array will then be overgrown with silicon, and the gate structures will be laid down by electron beam lithography (EBL). This approach, although more difficult to implement, could in principle allow us to build a Kane-type computer with the required precision. The bottom-up approach is not discussed here but is described in detail in the article "Fabricating a Qubit Array with a Scanning Tunneling Microscope" on page 302.

Top-Down Approach for Creating a Two-Donor Device. A host of issues surrounds the operation and readout of the nuclear-spin quantum computer. A key concern involves the transfer of the electron from the target to the readout atom during readout. The two electrons in the D^- state are not bound very strongly to the phosphorous atom, and the electron may be lost during the transfer. The initial phosphorus-phosphorus (P-P) system would transform into a P-P$^+$ system. If that is the case, we may need to use extra electrodes in order to create a deeper potential that will confine the electron or to employ a different readout atom (such as tellurium) that has a more strongly bound two-electron state.


(a) Two-Donor Test Device

(b) Top View Schematic

(c) Two-Donor Test Device

(d) Fabricated Test Device

(e) Magnified View

(f) J-Gate

Figure 8. Fabrication of a Two-Donor Device
(a) For proof-of-principle experiments involving the transfer of an electron between two phosphorus donors, it is not necessary to configure A-gates above the donors. Instead, we are configuring the three-gate device shown in side view in this illustration. (b) The top view schematic shows that the donors will be implanted between the A- and J-gates (red circles). The single horizontal line at the top of the gates represents an opening in a polymer resist layer. Ions can enter the substrate only where the line crosses the gates. The schematic also shows representative SET devices located near the implantation sites. (c) AFM image of an actual device prior to implantation. The narrow horizontal line near the end of the gates is a 20-nm-wide opening in the resist. (d) An SEM image of a fabricated metal-gate array is shown. The large metal structures on either side of the gates are aluminum electrodes used to detect the impact of ions during implantation. (e) This magnified view shows the central A-, J-, and A-gates. We have fabricated gate arrays with J-gate widths of less than 15 nm and gate separations down to 30 nm. The image in (f) shows a J-gate made from a titanium/gold alloy that is only 12 nm wide.

Our current goal for the top-down approach is to produce a device that can be used to study the controlled electron transfer between two donors. We intend to ionize one of the two phosphorus atoms and then study the coherent transfer of the remaining electron between the two donor atoms in the P-P$^+$ system. For this purpose, we have relaxed the stringent constraints of the Kane computer architecture and designed the simple device shown in Figure 8. It consists of two A-gates separated by a single J-gate. Two phosphorous atoms will be implanted between the A- and J-gates, an arrangement that is sufficient for charge transfer experiments.

Device fabrication starts with a wafer that is already topped with a barrier layer of SiO_2. To deposit metal A- and J-gates on the surface

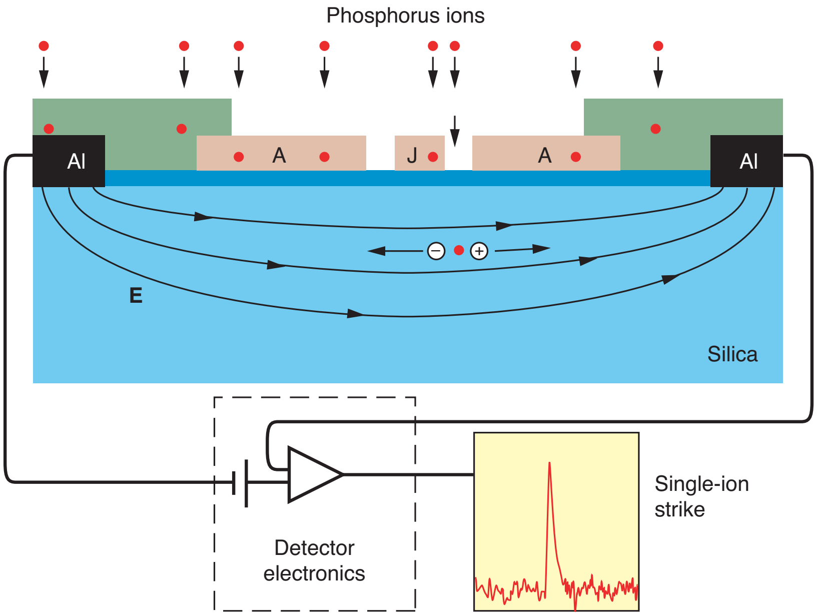


Figure 9. Detection of a Single-Ion Impact

We can detect in situ the impact of a single phosphorous ion in the silicon wafer. Aluminum electrodes are deposited on either side of the implantation site. An electric field applied between the two electrodes is used to separate electron-hole pairs that are created by ion impact. Each type of charge carrier migrates to its respective electrode and produces a current pulse that is detected by an external circuit. The implantation is halted after two such pulses have been detected.

before ion implantation, we use EBL techniques. A new layer of resist is created that covers the entire surface, including the gates. A second EBL exposure then patterns a thin line across the gates. This pattern is developed so that two tiny channels, each approximately 15 × 30 nanometers, are created on each side of the J-gate. The channels extend down to clean SiO_2 and define the implantation sites.

Next, we bombard the wafer with phosphorous ions. Although most of the ions are stopped in the mask, some go through the channels, strike the wafer, and get implanted about 10 nanometers below the Si/SiO_2 interface. After implantation, the device is heated to between 900°C and 950°C to anneal any damage to the silicon lattice. As a final step, we lay down the SETs. Creating the SETs after the anneal (instead of making them in the same step as the control gates) protects their fragile tunnel barriers, which would likely be degraded should they be submitted to temperatures above 900°C.

Because we want only one phosphorus atom per implantation site, the key to this entire process is the ability to detect a single ion after it has struck the silicon. And it is the properties of the silicon itself that help us fulfill this task. The energetic phosphorous ion produces a cascade of electron-hole pairs as it slows down and comes to a stop in the silicon matrix. Those charge species can be separated by an applied electric field, accelerated, and detected as a current pulse in an external circuit (see Figure 9). Voltage applied to surface electrodes straddling the implantation sites produces the field and transmits the pulses. The intrinsic silicon substrate makes this in situ particle-detection system highly efficient because the accelerating electric field extends fully between the two electrodes. We have demonstrated detection efficiencies of over 99 percent. Unfortunately, we cannot tell where the ion falls, and as there are two holes, there is only a 50 percent chance of creating a two-donor device with one donor in each hole. Although there are ways to improve those odds (by masking all but one hole with a specialized AFM cantilever, for example), in the short term, a 50 percent success rate is acceptable.

So far, no one has come close to seeing the transfer of a single electron between two precisely located, nanofabricated donor atoms. We hope to do so with a device similar to the one described above by September or October of 2002. We would then be in a position to study the coherent transfer of the electron between two donor atoms and possibly obtain information on decoherence mechanisms of relevance to spin readout.

Unlike the simple test device shown in Figure 8, an ideally configured device would have the A-gates directly above the phosphorous donors. We have designed a process to fabricate such a device (McKinnon et al. 2001). A multilayer electron-beam-sensitive resist is deposited on top of a SiO_2-coated silicon wafer, the resist is partially developed, and a linear array of ion-implantation channels is patterned in the resist with EBL techniques. The wafer is bombarded with phosphorous ions. The resist is then fully developed, and triple-angle metal evaporation is used to deposit the metal gate array and the SETs on the SiO_2 surface. Because only one mask is used to define the location of both the ion implantation sites and the gates, the A-gates are automatically registered over the qubits.

Figure 10 shows a six-donor test

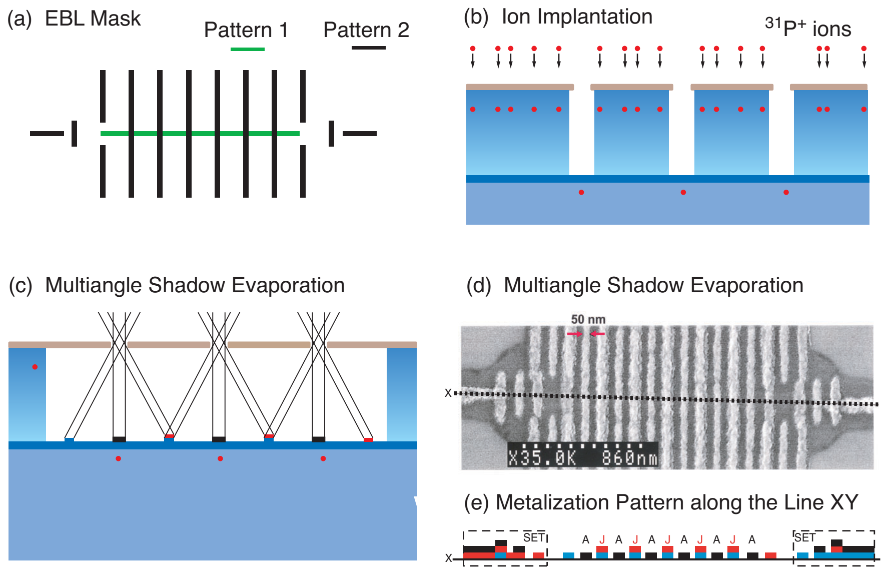


Figure 10. Creating a Multidonor Device

We have developed a process for creating gates directly over implanted ions. (a) In separate steps, patterns 1 and 2 of this EBL mask are partially developed in a trilayer resist that coats the substrate. The resist then sustains a series of lines with tiny channels that extend down to the substrate, where patterns 1 and 2 intersect. (b) A cross section of the resist after partial development shows the channels. The wafer is now bombarded with phosphorus ions. Some ions make it to the silicon surface and are implanted 5–10 nm below the Si/SiO_2 interface. (c) The resist is fully developed and material is removed, leaving behind a large cavity between the SiO_2 surface and the self-supporting top layer. Triple-angle shadow evaporation is then used to lay down an array of metal gates on the SiO_2 surface. (d) This photo shows a potential six-donor device with two readout SETs. No ions were implanted into this device. (e) A schematic side view of the device reveals the metalization pattern that results from the triple-angle shadow evaporation process.

device made by triple-angle evaporation. No ions were implanted into this device, and as can be seen in the figure, the triple-angle process does not yet result in gates that are sufficiently narrow to allow implementing the Kane scheme. We also have to address the problem of maintaining the integrity of the SET through the high-temperature anneal. However, the fundamental idea is robust.

The precision of the donor arrays produced by the top-down approach will be limited by straggling, which is inherent to ion implantation, and by the diffusion of dopants during the annealing step. Recent calculations indicate that small irregularities in the ion array could impair the operation of the quantum computer. That is why we are also pursuing the bottom-up fabrication approach, which might lead to a device with a very regular, well-characterized donor array.

Concluding Remarks

Phosphorous in silicon is a very clean, well-understood solid-state system. In its turn, NMR is a very well understood nuclear-spin manipulation technique. Performing NMR on a silicon chip implanted with phosphorus can therefore make for a very powerful quantum computer.

But the creation of a silicon-based solid-state computer presents such an enormous technical challenge that we must explore several strategies for building and implementing almost every aspect of the device. Hence, we investigate both SETs and magnetic resonance force microscopy as a means to read out the qubit state. Similarly, we have pursued two complementary fabrication strategies: the top-down process, which uses industrial production techniques, such as ion implantation and EBL, to produce a few-qubit device, and the bottom-up process, which involves advanced STM techniques and conventional molecular-beam epitaxy. Although the bottom-up approach is less suited to high-throughput production, it has the potential of leading to large, highly regular qubit arrays. We have made significant progress along all these parallel development paths.

Currently, scaling up a solid-state computer to over a million qubits is a goal that appears so distant as to be nearly out of sight. Yet less than fifty years ago, computer companies attempting to reduce the size of their machines were just becoming aware of a new strategy known as integrated circuits. Those early chips were crude and contained but a few transistors, but from them, evolved the modern, densely packed integrated circuits of today. Like those early chips, the quantum devices developed so far are rudimentary. No doubt, the challenges we face in building a real silicon-based quantum computer are significant, but our initial results offer hope that large-scale quantum computing may one day be realized. ■

Acknowledgments

Building a silicon-based solid-state quantum computer is a difficult and challenging project that cannot be achieved without the efforts of a large research team. At the University of New South Wales, Tilo Buehler, Rolf Brenner, David Reilly, Bob Starrett, and Dave Barber have worked in the development of SET and rf SET readout schemes. Similarly, important contributions to the top-down fabrication approach have been made by Fay Stanley, Eric Gauja, Nancy Lumpkin, Rita McKinnon, Mladen Mitic, Linda Macks, and Victor Chan. At the University of Melbourne, Cameron Wellard has supported the device modeling; Changyi Yang, Paul Spizzirri, and Robert Short have worked on single-ion implantation technologies; and Jeff McCallum and Steven Prawer have contributed in a range of areas. The Magnetic Resonance Force Microscope research would not have been possible without the contributions of Denis Pelekhov, Michael Roukes, Andreas Suter, and Zhenyong Zhang.

Further Reading

Feher, G. 1959. Electron Spin Resonance Experiments on Donors in Silicon. *Phys. Rev.* **114** (5): 1219.

Goan, H., and G. Millburn. 2000 (February). Silicon-Based Electron-Mediated Nuclear Spin Quantum Computer. CQCT, University of Queensland internal report.

Gordon, J. P., and K. D., Bowers. 1958. Microwave Spin Echoes from Donor Electrons in Silicon. *Phys. Rev. Lett.* **1**: 368.

Honig, A., and E. Stupp. 1960. Electron Spin-Lattice Relaxation in Phosphorus-Doped Silicon. *Phys. Rev*. **117** (1): 69.

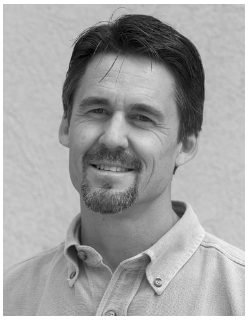
Kane, B. E. 1998. A Silicon-Based Nuclear Spin Quantum Computer. *Nature* **393**: 133.

McKinnon, R. P., F. E. Stanley, T. M. Buehler, E. Gauja, K. Peceros, L. D. Macks, M. Mitic et al. 2002. Nanofabrication Processes for Single-Ion Implantation of Silicon Quantum Computer Devices. (To be published in *Smart Mater. Structures*.)

Mozyrsky, D, Sh. Kogan, and G. P. Berman. 2001. Time Scales of Phonon Induced Decoherence of Semiconductor Spin Qubits. [Online]: http://eprints.lanl.gov (cond-mat/0112135).

Schoelkopf, R. J., P. Wahlgren, A. A. Kozhevnikov, P. Delsing, and D. E. Prober, 1998. The Radio-Frequency Single-Electron Transistor (RF-SET): A Fast and Ultra-sensitive Electrometer. *Science*, **280**: 1238.

Robert G. Clark received Bachelor of Science and Ph.D. degrees in physics from the University of New South Wales in Sydney, Australia, in 1973 and 1983, respectively. Between 1984 and 1990, Bob was a lecturer in physics at the University of Oxford, England, and research group head at Clarendon Laboratory at Oxford. During this period, he received a number of awards and distinctions, including the Wolfson Award for prestigious research (1988). In 1990, Bob became professor of experimental physics at the University of New South Wales. He is now heading the Center for Quantum Computer Technology in Sydney, Australia, and is leading the Australian effort for developing a silicon-based solid-state quantum computer.



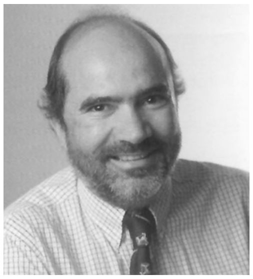
P. Chris Hammel received his B.A. from the University of California at San Diego and his Ph.D. from Cornell University (1984). Following a postdoctoral appointment at the Massachusetts Institute of Technology, Chris came to Los Alamos National Laboratory on a J. Robert Oppenheimer Fellowship. He became a staff member at Los Alamos in 1989. He is a Fellow of Los Alamos National Laboratory and was awarded the Los Alamos Fellows Prize in 1995. He is also a Fellow of the American Physical Society. In June, 2002, Chris joined the Physics Department at the Ohio State University as an Ohio Eminent Scholar. In addition to his current research, which is focused on the development of an ultrasensitive magnetic resonance force microscope, he is actively engaged in magnetic resonance studies of correlated electron systems including high-temperature cuprates and heavy fermions.

Andrew Dzurak is a program manager at the Centre for Quantum Computer Technology and Associate Professor within the School of Electrical Engineering at the University of New South Wales, Sydney. He has collaborated with researchers at Los Alamos since 1996. Andrew has 15 years of experience in the fabrication and cryogenic electrical measurement of quantum effects in semiconductor nanostructures, beginning with a Ph.D. at the University of Cambridge (the United Kingdom). Andrew moved to UNSW to take up an Australian Postdoctoral Fellowship and has worked closely with Robert Clark to establish the Semiconductor Nanofabrication Facility, of which he has been Deputy Director since 1995. Within the Centre, Andrew manages research in integrated devices for the control and readout of phosphorous qubits in silicon.

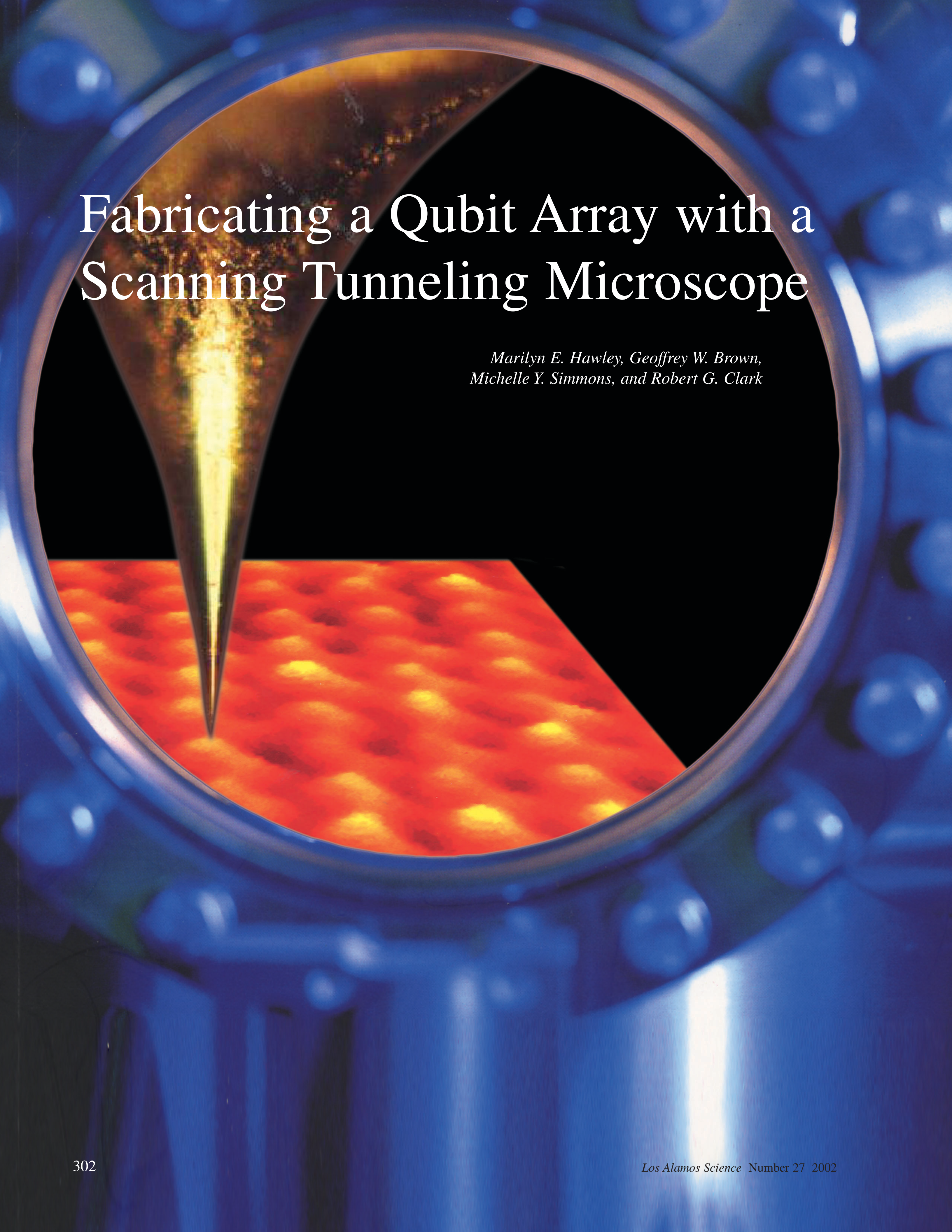
Alex Hamilton received his Ph.D. from the University of Cambridge, the United Kingdom, in 1993. He is a senior lecturer in physics in the School of Physics at the University of New South Wales, as well as manager for the Quantum Measurement Program at the Centre for Quantum Computer Technology. He is interested in the study of the fundamental properties of submicron electronic devices, with particular interest in the regime in which strong interactions between adjacent devices influence the behavior of the complete system.

Llloyd Hollenberg is a senior lecturer at the University of Melbourne, and since 2001 has been manager of the Device Modeling Program at the Centre for Quantum Computer Technology. He received his Ph.D. from the University of Melbourne in 1989. His interests include the general aspects of quantum many body systems, and in particular field theoretic and spin systems. At present, his work focuses on the physics of the Kane quantum computer and on quantum algorithms.

David N. Jamieson is an associate professor and reader at the University of Melbourne, Australia. He received his Ph.D. from the University of Melbourne, and has been Director of the Microanalytical Research Centre (MARC) of the School of Physics, University of Melbourne since 1996. In 2000, he joined the Centre for Quantum Computer Technology as the manager of the Ion Beam Program. He aims to adapt the technology of single ion detection and control to the construction of phosphorous arrays in silicon.



Christopher I. Pakes received his Ph.D. in physics in 1999 from the College of Physics, Birmingham, the United Kingdom. He has studied metrological applications of nanotechnology and spin-polarized properties of oxide functional materials, using ultrahigh vacuum, low temperature STM and dc SQUIDs. In 2000, Chris joined the Center for Quantum Computer Technology and was appointed manager of the Atomic Level Manipulation and Imaging Program in 2001.



Fabricating a Qubit Array with a Scanning Tunneling Microscope

Marilyn E. Hawley, Geoffrey W. Brown,
Michelle Y. Simmons, and Robert G. Clark

The Australian Centre for Quantum Computer Technology and Los Alamos National Laboratory are working together to answer the question, "Can the solid-state quantum computer (SSQC) proposed by Bruce Kane (1998) be built?" Illustrated in Figure 1, the architecture put forward by Kane requires a linear array of phosphorus atoms (nuclear spin 1/2) inside an isotopically pure silicon-28 (spin 0) wafer. The spacing between the atoms needs to be about 20 nanometers, and the array will be located 5 to 20 nanometers beneath the silicon surface. An array of metal electrodes, isolated from the silicon by a thin insulating layer of silicon dioxide (SiO_2), will sit above the qubit array and needs to be precisely registered to it. Because the array is so small and because the silicon overlayer must be nearly free of impurities and crystalline defects for the computer to operate properly, we must achieve unprecedented control of the fabrication process.

Our efforts to build the SSQC focus on a novel "bottom-up" fabrication approach. Starting with a clean silicon surface, we will build each layer of the device in succession, first creating the phosphorus array and embedding it in the surface, then growing the silicon overlayer, the SiO_2 insulating layer, and finally laying down the metal electrodes. (We are also pursuing a "top-down" fabrication approach, which along with information about the operating principles of the computer, is described in the article "Toward a Silicon-Based Nuclear-Spin Quantum Computer" on page 284.)

The scanning tunneling microscope (STM) plays a central role in the bottom-up approach, serving as both a fabrication and electrical characterization tool. To create the phosphorus array, we employ STM-based hydrogen lithography, developed by Joseph Lyding's group at the University of Illinois at Urbana-Champaign (Lyding et al. 1994). Immediately following the array fabrication step, the silicon overlayer will be grown by molecular beam epitaxy to encapsulate the array. Our STMs have variable temperature control so that we can anneal the overlayer in situ, and thus be in a position to study the stability of the phosphorus array during silicon overgrowth. We can also identify potential defects and impurities that could impair computer operation. Once the thin SiO_2 layer is grown, we will create the metal-gate array using state-of-the-art electron beam lithography (EBL) technology.

In this article, we summarize our progress in building the phosphorus array, overgrowing the silicon layer, and checking whether the latter step alters the array. To convey the central role of the STM in building the SSQC, we start by presenting the principles that make it such a powerful fabrication and characterization tool.

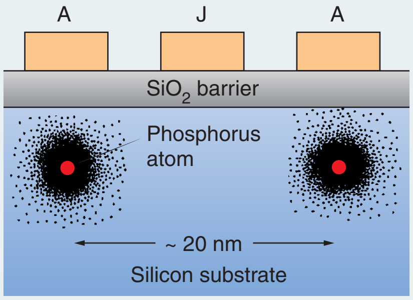


Figure 1. Kane's Architecture for a Quantum Computer
In Kane's concept of a silicon-based quantum computer, the qubits are phosphorus atoms embedded in an isotopically pure ^{28}Si crystal at a distance of about 20 nm from each other. Above the silicon, there is an insulating barrier of SiO_2, and above that barrier, metallic gate electrodes. The A-gates help manipulate the individual qubits whereas the J-gates control the interaction between neighboring qubits.

Scanning Tunneling Microscopy

The STM probes the surface of a sample by inducing electrons to tunnel between the surface and the tip. As illustrated in Figure 2, an extremely sharp metallic tip (with radius of curvature *R* that is typically about 10 nanometers) is brought to within a few angstroms of a sample's surface. The thin vacuum region separating the tip and the sample forms a potential barrier, and a bias voltage between the tip and the sample causes more electrons to tunnel through the barrier from occupied energy states to unoccupied ones. To a first approximation, the tunneling current at a point on the surface is proportional to the local electron density of states (LDOS) in the sample. By measuring the tunneling current as a function of position, we can obtain an extremely localized map of the electronic structure of the sample's surface.

The tip is attached to a piezoelectric scanning device, which moves it over the surface of the sample in a raster pattern. An image of the surface is thus obtained. In practice, we use a feedback loop to adjust the tip height and keep the tunneling current constant as the tip moves. (Scanning in this "constant-current" mode prevents the tip from crashing into protrusions, such as surface steps.) The resulting map of tip heights versus position can be used to construct an image of the surface that shows contours of constant LDOS. On many surfaces, this

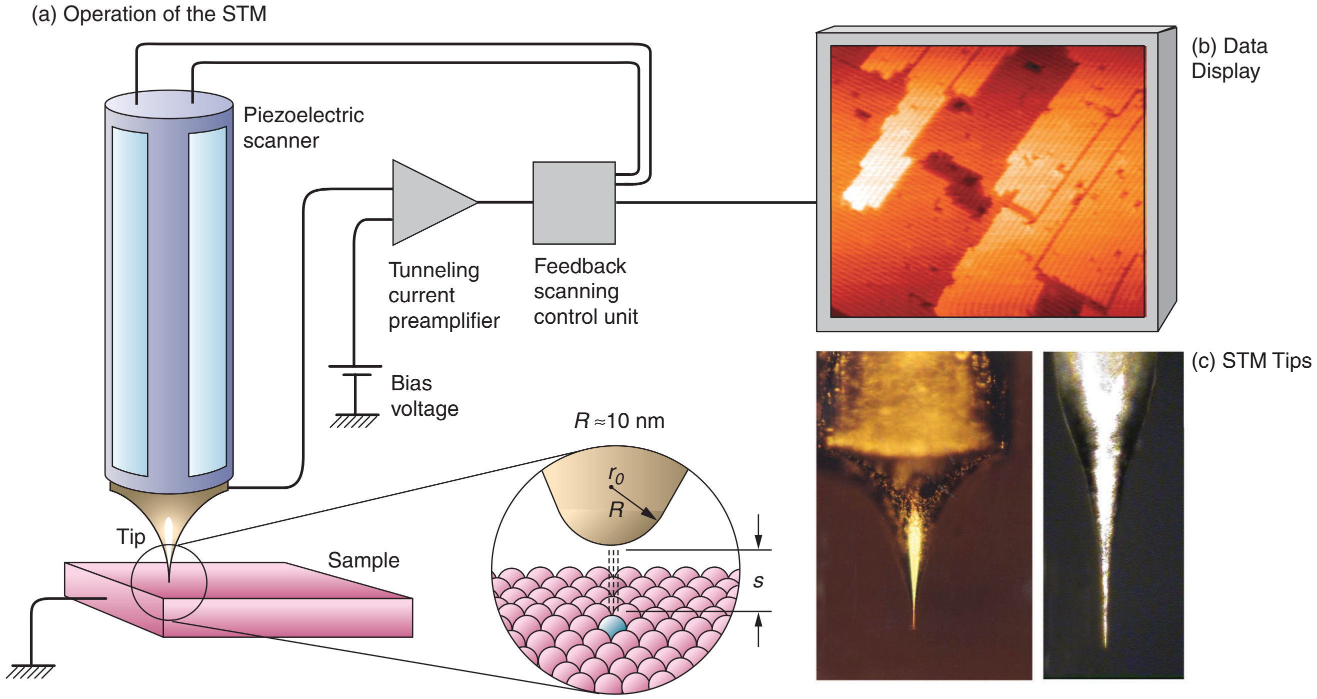


Figure 2. The Scanning Tunneling Microscope (STM)

(a) This schematic diagram illustrates the steps in the operation of an STM. An extremely sharp tip is held within a few tenths of a nanometer of a sample surface. A bias voltage *V* applied between the tip and the sample causes electrons to tunnel between the two. The tunneling current is monitored with a feedback loop, which keeps the current constant by varying the gap width *s* between the tip and the sample. The gap width is then proportional to the sample's local density of states (LDOS). The tip moves in a raster pattern laterally over the surface. A plot of the tunneling current versus position is a map of the sample's LDOS. One such map is shown in (b). In many cases it is equivalent to a map of atomic positions. (c) These optical micrographs show two STM tips: One is made of tungsten (left) and the other, of etched 90% Pt–10% Ir alloy. Each tip has a radius of curvature of about 10 nm.

contour map is equivalent to a map of the atomic positions.

The electronic-energy diagrams of the tunneling process, shown in Figure 3, help to explain the technique's atomic resolution, as well as the subtleties of the information obtained. The applied bias voltage defines the energy offset, or energy "window," between the Fermi levels of the tip and the sample. Any electrons that have energies within that window contribute to the net tunneling current.

In 1985, shortly after the development of the STM, Jerry Tersoff and Donald Hamann described the tunneling mathematically, by applying Bardeen's tunneling theory (1961) to the tip-sample system. By assuming a low temperature, a small bias voltage *V*, and a featureless tip (one in which the electron density of states is constant), they showed that the tunneling current could be written as

$$I_t \propto \sum_{v} \left|\Psi_v(r_0)\right|^2 \delta\left(eV - E_{\mathrm{F}}\right) \quad . \qquad (1)$$

Here, Ψ_v are the sample's wave functions whose energy eV above the Fermi level E_{F} is evaluated at the point r_0 on the tip—see Figure 2(a). The sum over the probability densities from all such wave functions is the LDOS of the sample directly below the tip, so that in the approximation of Equation (1), the tunneling current is indeed proportional to the sample's LDOS.

The spatial resolution of an STM image is extremely high (approximately 0.01 angstrom) in the direction perpendicular to the surface. That is so because the tunneling probability T decreases exponentially with the separation s between the tip and the sample. The Wentzel-Kramers-Brillouin (WKB) approximation for the tunneling probability through the type of potential barrier shown in Figure 3 (a trapezoidal barrier between planar metal electrodes) yields

$$T = e^{-2\kappa s} \quad , \qquad (2)$$

where κ, the inverse decay constant in the potential barrier, is given by

$$\kappa = \sqrt{\frac{2m_e}{\hbar^2}\left(\frac{\phi_{\mathrm{t}} + \phi_{\mathrm{s}} - eV}{2} - E\right)} \quad . \qquad (3)$$

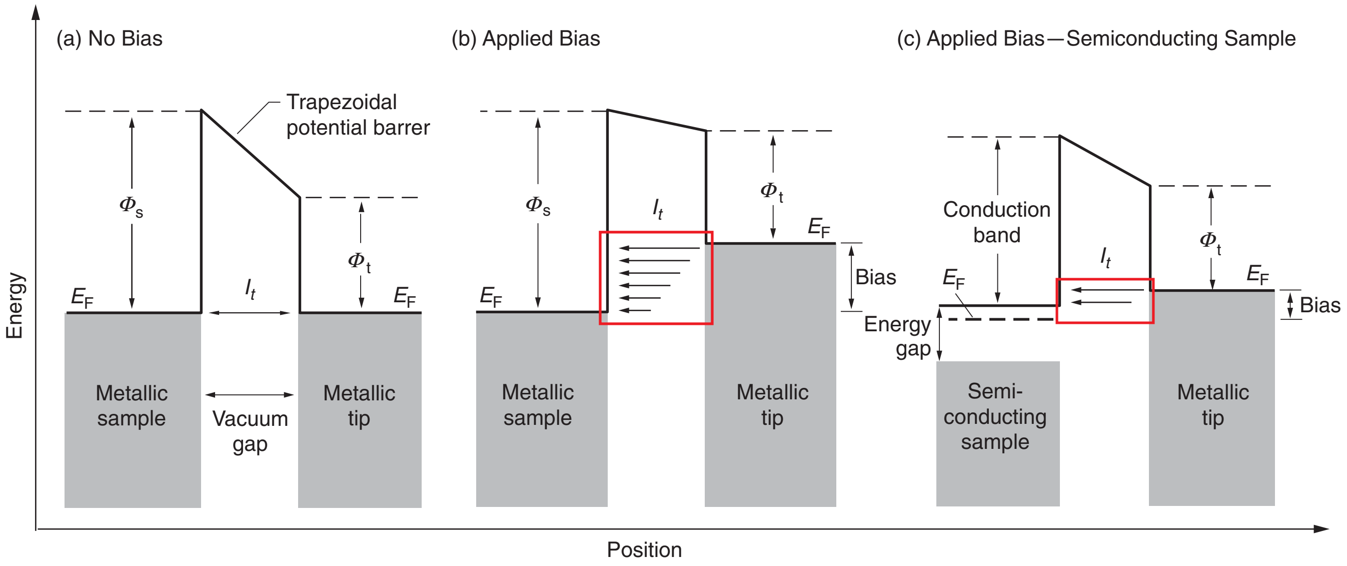


Figure 3. Tunneling Process for Metallic and Semiconducting Samples

This schematic electronic-energy level diagram helps illustrate basic STM concepts. Filled electronic states are below the Fermi level E_F whereas empty states are above. A potential barrier is created by the vacuum gap between the tip and sample. If the width of the barrier is so narrow that the electron wave functions of the tip and sample overlap, then electrons can tunnel to empty states in either the tip or the sample. (a) When the two Fermi levels are equal (because the tip and sample are connected to a common ground) there is no net current flow. (b) When a bias voltage is applied, the Fermi levels of the two materials become unequal, and the difference defines an energy window (red box). In the case shown, the bias voltage raises the Fermi level in the metal tip relative to that in the metal sample. Electrons in filled states within the energy window can tunnel from the tip, through the potential barrier, into the sample's empty states. The arrows of decreasing size indicate that the tunneling probability is highest for electrons at the Fermi level of the tip and decreases as the electron energy decreases. (c) No states are available in the energy gap between filled states and the conduction band of a semiconductor sample. Electrons can only tunnel into empty states in the conduction band.

In Equation (3), m_e is the free-electron mass in vacuum, and $\hbar$ is the reduced Planck constant. The variables are the work functions[1] of the tip and the sample, ϕ_t and ϕ_s, respectively, the electron kinetic energy normal to the barrier E (measured relative to the tip's Fermi level), and the bias voltage V applied to the sample.

Given nominal values for the parameters in Equation (3) (for example, $\phi_t \approx \phi_s = 3$–$6$ electron volts, $E \approx 0.025$ electron volts, and $V = 1$–2 volts), the decay constant κ is of the order of 0.1 nanometer^{-1}. A change of 0.1 nanometer in the spacing between the tip and the sample alters the tunneling probability by $e^2 = 7.4$. Thus, a topographic resolution of the order of 0.001 nanometer in the direction perpendicular to the surface requires only a 2 percent precision in the measurement of the tunneling current. With carefully designed, low-noise electronics, that precision is easily achieved—even for a tunneling current of 100 picoamperes.

The resolution parallel to the surface is also atomic—on the order of 0.1 nanometer—for much the same reason: The extreme sensitivity of the tunneling current to the gap width ensures that essentially the entire tunneling current arises from a single atom or a small cluster of atoms at the very end of the tip (those atom(s) closest to the sample). On a clean, well-formed tip with a small radius of curvature, atoms or clusters that are laterally displaced from the end are also farther from the sample and do not contribute a significant number of electrons to the tunneling current. Thus, there is very little lateral spread associated with the signal.

A more-detailed look at the origin of the tunneling current will shed additional light on the information contained in the STM image. Equation (1) can be rewritten to account for both a finite energy window for tunneling and a more-complex electronic structure of the tip as follows (Selloni 1985):

$$I_t \propto \int_{-eV}^{0} \rho_t(E)\ \rho_s(E+eV)\ T(E,eV)\ dE\,, \quad (4)$$

where ρ_t and ρ_s are the tip and sample LDOS, respectively, T is the tunneling probability between the tip and the

---

[1] The work function ϕ is the energy needed to remove an electron, whose energy is at the Fermi level, from the sample.

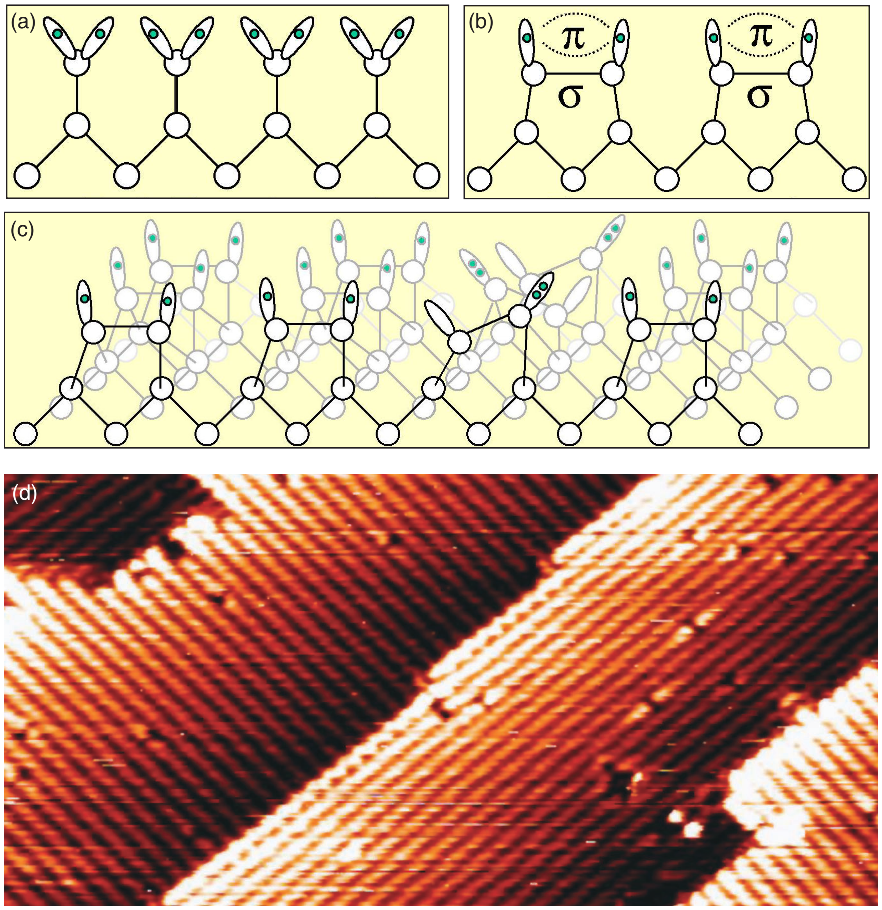


Figure 4. The Si(100)-(2 × 1) Surface
(a) Schematic view of a row of atoms in a bulk-terminated Si(100) surface. This configuration is energetically unfavorable because every atom has two singly occupied dangling bonds. (b) The bulk-terminated surface can reconstruct into the so-called Si(100)-(2 × 1) surface. The dangling bonds of neighboring atoms join to form σ-bonded dimers, and the remaining protruding bonds become weakly π-bonded. (c) This view of the reconstructed surface shows several rows of dimers. In the third row from the left, the dimers are pinned in the buckled configuration (see text). (d) A filled-state image of a Si(100)-(2 × 1) surface (10 nm × 5 nm) showing several monolayers. The bright lines making up each "terrace" are the dimer rows, which rotate by 90° with each successive layer. The defect density of an average sample is approximately 5%. The wavy line along the upper edge of the central terrace is due to buckled dimers.

sample, and the dependence of the tunneling current on r_0 has been suppressed for simplicity.

The expression in Equation (4) emphasizes that the properties of both the tip and the sample contribute to the tunneling current. Therefore, one needs to have substantial background information about both in order to interpret an STM image. For example, numerous geometric and electronic effects go into the LDOS function ρ_s, including the electronic structure at the surface, the band structure of the material, the presence of dangling bonds at the surface or bulk resonances, the number and orientation of back bonds, and so forth. In addition, we need information about the electronic structure of any adatoms or contaminants (such as oxygen, carbon, carbon dioxide, nitrogen, and others) that might be present. Through ρ_t, the appearance of an image is also closely related to the electronic structure of the tip. Finally, of particular importance is the fact that the electron density is not always centered about the cores of the atoms in the material. Several of these considerations arise in our work on the SSQC and will be discussed later.

One powerful technique that can be used to help us interpret images is to change the direction (sign) of the bias voltage. By doing so, we cause the tunneling current to reverse its direction. If the tip is biased to have a higher Fermi level, then current flows from the tip to the empty states in the sample. If the bias is reversed, so that the sample has a higher Fermi level, then the electrons from the sample's filled states flow into the empty states in the tip. We therefore have a means to obtain information about the density of both the empty and filled states of the sample. The differences between the two STM images help us sort out electronic effects from structural information and to distinguish among features that appear identical when only one bias direction is used.

Despite the intricacies involved, we can interpret an STM image quite accurately when all the available information is taken into account. That is why STM imaging is continuing to produce significant results in surface science.

Preparing Silicon(100) Surfaces

The bottom-up fabrication approach begins by preparing a flat (100)-oriented silicon surface. Technologically, this is one of the most important semiconductor surfaces. For our purposes, it is relatively easy to prepare, can be patterned by

STM-based hydrogen lithography, and is well suited for the subsequent overgrowth of crystalline silicon layers. Although the (100) surface has been studied by STM and other methods for over 15 years, we are uncovering new details important to constructing the type of atomic-scale electronic structures needed in the Kane solid-state quantum computer and other quantum devices.

Figures 4(a) and 4(b) show how the bulk-terminated (100) surface, every atom of which has two dangling bonds, reconstructs in a manner that lowers the surface energy. Electrons from two neighboring silicon atoms form a σ-bond, so the resulting silicon-silicon dimer has only two dangling bonds. These bonds form a weak π-bond to further reduce the surface energy. The π-bond can easily be broken by chemically active species, such as hydrogen, which adsorb on the surface.

The reconstructed surface, commonly referred to as the Si(100)-(2 × 1) surface (a designation that derives from the corresponding electron-diffraction pattern), takes on the appearance of a neatly plowed field, with rows of dimers aligned parallel to each other, as seen in Figures 4(c) and 4(d). In filled-state STM images taken at room temperature, most of the dimers appear as symmetric bean shapes. In reality, the dimers are tilted, or buckled, and are flipping back and forth between buckled configurations very rapidly—refer to Figure 4(c). The oscillation takes place too quickly to be imaged with an STM. Therefore, in general, an average configuration is observed. Near defects or step edges, however, the dimer can be pinned in an asymmetric position and imaged. Such an image can be seen at several locations in Figure 4(d), where neighboring dimers are seen to buckle in alternate directions.

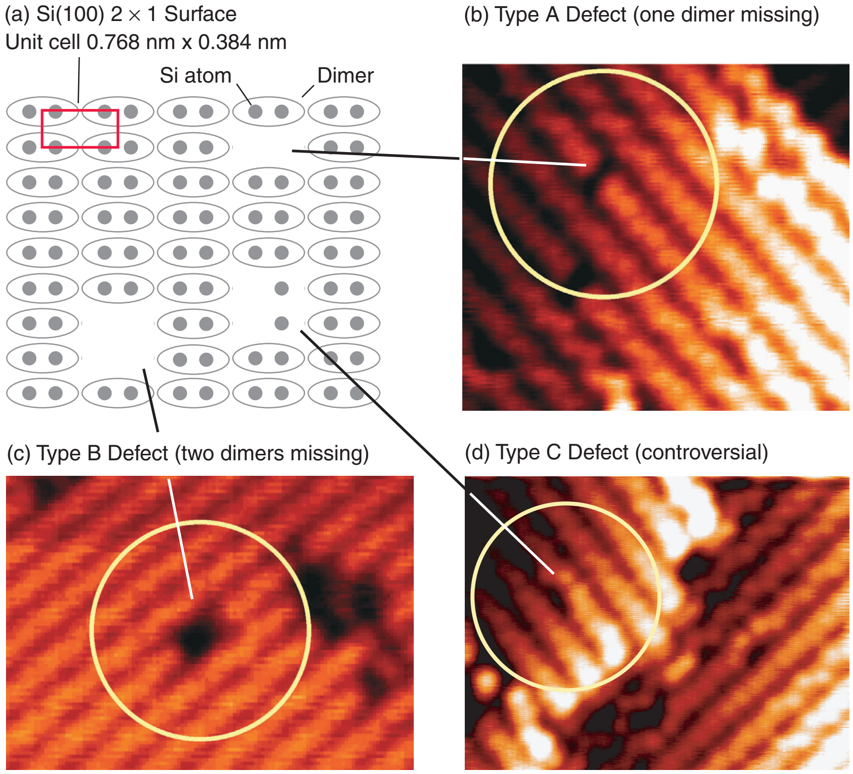

(a) Si(100) 2 × 1 Surface
Unit cell 0.768 nm x 0.384 nm
Si atom
Dimer
(b) Type A Defect (one dimer missing)
(c) Type B Defect (two dimers missing)
(d) Type C Defect (controversial)

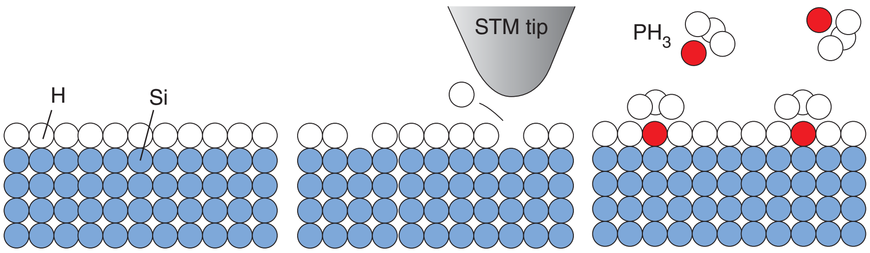
Figure 5. Defects in the Si(100)-(2 × 1) Surface

(a) This cartoon of the Si(100)-(2 × 1) surface illustrates type A, B, and C defects. (The structure of a type C defect is controversial. See text for a further discussion.) The 0.768 nm × 0.384 nm unit cell (the values are for the dimer spacings along the rows and between the rows, respectively) is also shown. (b), (c), and (d) show filled-state images of type A, B, and C defects, respectively.

Surface preparation begins with degassing the sample and its holder by holding them at an elevated temperature for several hours. (Because the surface is reactive, this step and those that follow are carried out in ultrahigh vacuum.) The sample is flash-heated to a temperature of 1250°C and cooled under conditions that allow the surface silicon atoms to form a well-ordered Si(100)-(2 × 1) surface. But the difficulty in precisely controlling the annealing process and the inability to cut the starting substrate exactly on axis result in a surface typically consisting of several terraces of simple atomic planes. On a given terrace, all the dimer rows run in the same direction, whereas the in-plane orientation of the dimer rows rotates by 90° from one terrace to the next. The terrace edges terminate smoothly or roughly, depending on whether the dimer rows for that terrace run parallel or perpendicular to the terrace edges, respectively.

Even a freshly prepared Si(100)-(2 × 1) surface will contain defects. The main types observed in STM images, illustrated in Figure 5, are type A defects, in which a single silicon dimer is missing, type B, in which two adjacent dimers in a row are missing, and type C, whose makeup is still controversial. Type C defects could be the result of a subsurface vacancy, or else consist of two missing silicon atoms from adjacent dimers in a row. They could also be due to an adsorbed impurity, for example an absorbed water molecule (Chander et al. 1993). Although it is nearly impos-

(a) Monohydride Deposition (b) Hydrogen Desorption (c) PH_3 Dosing



(d) Heterodimer Formation (e) Hydrogen Desorption (f) Silicon Overgrowth

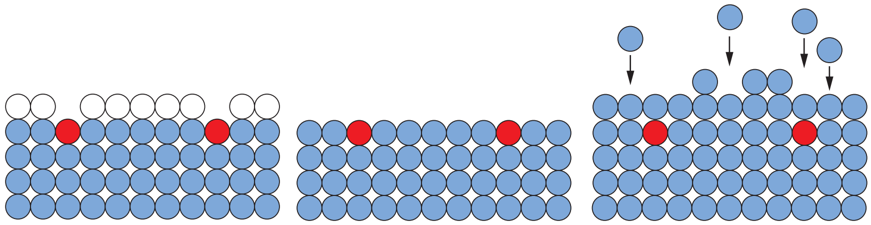


Figure 6. The Bottom-up Approach for Fabricating an Array of Phosphorus Qubits

(a) After preparing and cleaning a silicon surface, we dose it with hydrogen, which adsorbs as a monoatomic layer. (b) The STM tip selectively desorbs individual hydrogen atoms and exposes silicon at a set of regularly spaced sites that will define the qubit array. (c) PH_3 is introduced into the vacuum chamber. It bonds to the silicon only at the exposed sites. (d) A critical anneal is performed to incorporate the phosphorus atoms into the silicon surface, forming a P-Si heterodimer. (e) The hydrogen monolayer can be removed by further annealing at a slightly higher temperature (this step may not be necessary). (e) With molecular-beam epitaxy, the phosphorus array is buried under fresh layers of silicon.

sible to eliminate these defects during preparation, we can prepare surfaces with defect densities of less than a few percent by following careful vacuum practices in the STM chamber. The effect of defects on the operation of a quantum computer will be further discussed under the section "Qubits, Defects, and Dopants."

STM-Based Hydrogen Lithography

Once we have prepared a clean surface with a low defect density, we are ready to begin the array fabrication scheme. We use a resist technology analogous to the lithographic techniques used in conventional electronics manufacturing, the main difference being that the STM-based technology allows us to create features on the atomic scale.[2] The idea is illustrated in Figure 6.

The first step is to deposit a single layer of hydrogen atoms (the "resist") on the clean surface. In order to do so, we dissociate molecular hydrogen gas by passing it over a hot filament as it enters the STM vacuum chamber. The resulting hydrogen atoms are directed onto the heated sample surface, where they break the weak π-bond and adsorb to the surface by attaching to the very reactive dangling bonds. Provided the conditions are right, one hydrogen atom can covalently bond to each silicon atom, and the surface becomes coated with a uniform monohydride layer (see Figure 7).

The STM tip is then used as the lithographic patterning tool. Controlled-voltage pulses applied between the tip and the sample cause very small patches of the monohydride layer to vibrate and heat up and/or to become electronically excited. Individual hydrogen atoms are liberated, and as a result, the dangling bond of the underlying silicon atom becomes exposed. The tiny, atom-sized holes created by the STM are the only reactive sites on the otherwise unreactive monohydride layer. Interestingly, the holes created in the hydrogen layer appear as protrusions above the hydrogen-terminated surface. This is an example of electronic effects influencing the STM images. Whereas the hydrogen-terminated structures protrude farther into the vacuum than the dangling bonds, the energy of the dangling bonds is closer to the window between the Fermi levels of the tip and the sample. The dangling bonds, therefore, contribute more strongly to the tunneling current and appear "taller."

Next, we introduce high-purity phosphine (PH_3) gas directly into the ultrahigh-vacuum chamber of the microscope. The PH_3 is very reactive and adheres to the exposed dangling bond with a sticking coefficient of one. As seen in Figure 8, we can place single phosphorus-bearing molecules where necessary and thereby build an atomic-scale phosphorus array. The reacted sites appear taller than both the hydrogen-terminated sites and the unreacted dangling bonds. This effect is likely due to a combination of electronic and physical effects.

The next step is to stimulate the phosphorus atoms within the phosphine molecule (which is attached to the silicon atoms by a single bond) to incorporate into the top layer of the silicon surface and form a phosphorus-silicon heterodimer. In that structure, the phosphorus atom takes the place of one of the silicon atoms in the dimer and attaches to the remaining silicon surface through three strong covalent bonds. Formation of the heterodimer is a critical step because it secures the phosphorus atom in its patterned location and helps prevent its diffusion during subsequent processing steps.

---

[2] Scanning tunneling microscopy can be used directly to create atomically precise structures of metal atoms on metal surfaces. We are forced to adopt a lithography approach because the strong covalent bonds on the silicon surface prevent us from directly rearranging atoms using the STM.

Before studying the mechanism for incorporation through the hydrogen resist, we had to learn how to distinguish the postdosing phosphorus-related species from other features on the silicon surface because, to date, very few reports exist on the STM imaging of single phosphine molecules on silicon. We, therefore, conducted a series of experiments in which the clean Si(100)-(2 × 1) surface was subjected to various dosing conditions. Each time, the presence of phosphorus on the surfaces was confirmed by Auger electron spectroscopy. By examining both filled- and empty-state STM images, we found it was possible to distinguish between phosphine-related surface species and surface defects.

We then faced the challenge of phosphorus incorporation. It is well known that, at room temperature, phosphine adsorbs onto a clean Si(100)-(2 × 1) surface and quickly dissociates to form PH_2 and H. Subsequent heating of the surface to about 400ºC leads to the complete dissociation of PH_x ($x = 2–3$). We have demonstrated that, at these temperatures, the individual phosphorus atoms also incorporate into the surface and form the phosphorus-silicon heterodimer (see Figure 9). The hydrogen remains on the surface as a monohydride. Continued heating of the surface to higher temperatures will liberate the hydrogen. In this way, the surface is left clean, consisting of only silicon dimers and phosphorus-silicon heterodimers.

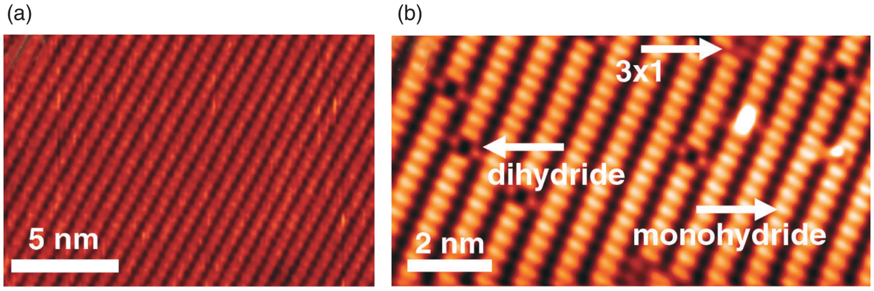
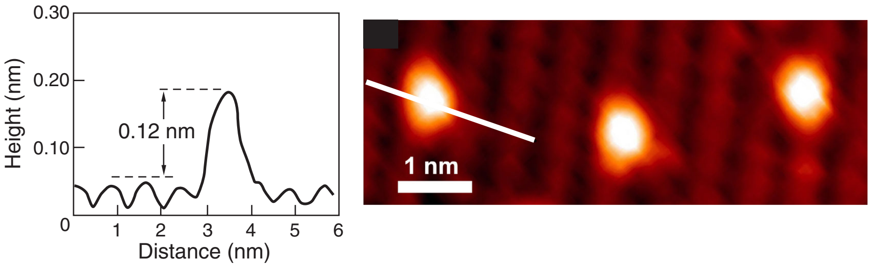


Figure 7. Creating the Hydrogen Resist
(a) This filled-state STM image is of a clean Si(100)-(2 × 1) surface with a very low defect density. (b) Shown here is a hydrogen-terminated Si(100)-(2 × 1) surface, which is almost entirely monohydride; that is, one hydrogen atom is bonded to each silicon atom. Several other structures are also apparent: dihydrides (two hydrogen atoms have bonded to a single silicon atom) and a 3 × 1 structure (three hydrogen atoms have bonded to one silicon atom).

(a) Desorption of Single Hydrogen Atoms in Monohydride Layer



(b) Adsorption of Phosphine in Desorption Sites

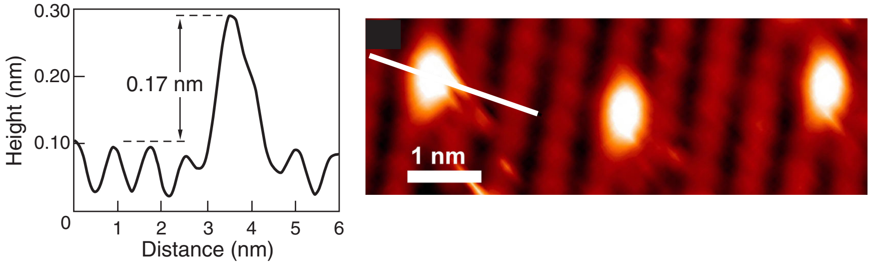


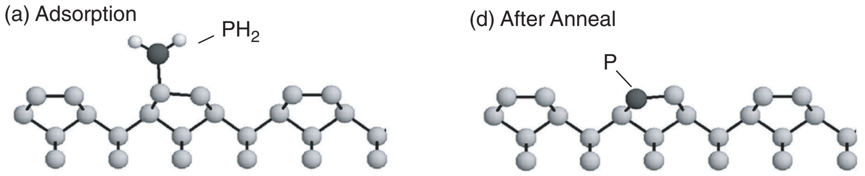
Figure 8. Adsorption of Single Phosphine Molecules
(a) This STM image (right) shows three desorption sites in a monohydride layer, and the graph (left) shows the line profile, taken along the indicated white line, of the leftmost site. The bright protrusion at each of the desorption sites is the signature of the single silicon dangling bond after desorption of just one hydrogen atom. (The sites appear brighter because their DOS are closer to the Fermi level, so they contribute more to the tunneling current.) (b) The same sites after dosing the surface with phosphine gas. The profile shows an increase of 0.05 nm in height (calibrated against an atomic step edge on the same surface), a reproducible increase that is observed at all adsorption sites. Given the information we gathered by scanning tunneling microscopy, our interpretation of the increase in height is that phosphine has adsorbed to the exposed sites.

Molecular Beam Epitaxy of Silicon

Subsequent steps for the fabrication of the SSQC call for growing a 50- to 200-angstrom-thick layer of crystalline silicon over the array of phosphorus atoms, depositing an insulating layer of SiO_2, and aligning gate electrodes to the now buried phosphorus array. High-quality crystalline, or epitaxial, growth of silicon on silicon is typically done at high temperatures.

However, it is known that at high temperatures, phosphorus atoms buried in silicon tend to diffuse upwards and pop up to the surface. Furthermore, we observed during our incorporation studies that, at temperatures of 650ºC and above, the phosphorus becomes mobile. It breaks from the heterodimer and begins to migrate about the surface until it meets another phosphorus atom. It then forms P_2 (or possibly P_4), which desorbs from the surface. Thus, the next significant question in the bottom-up approach is, "Can crystalline silicon be grown on either a clean or monohydrided surface at temperatures low enough to prevent the diffusion and segregation of phosphorus?"

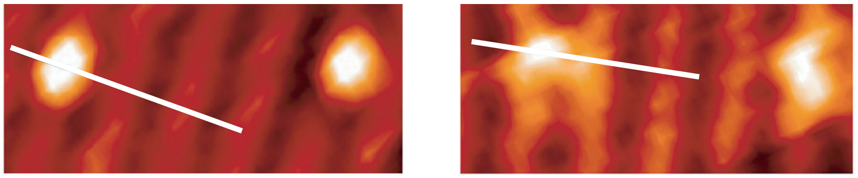
Taking into account results from the literature and our own experiments, we have adopted two parallel growth strategies. We begin both by annealing the sample directly after phosphine dosing, so that the phosphorus atoms become incorporated into the silicon surface and the hydrogen resist can desorb. We then need to encapsulate the phosphorus atoms under a few monolayers of silicon. In the first growth strategy, we will grow the encapsulation layer at room temperature. The resulting layer will have a high surface roughness with numerous silicon islands and require a subsequent annealing step for surface flattening. In the second strategy, we


(a) Adsorption

(d) After Anneal

(b) Two Adsorbed PH_3 Molecules

(e) Filled-State Heterodimer



(c) Line Profile after Adsorption

(f) Line Profile after Annealing

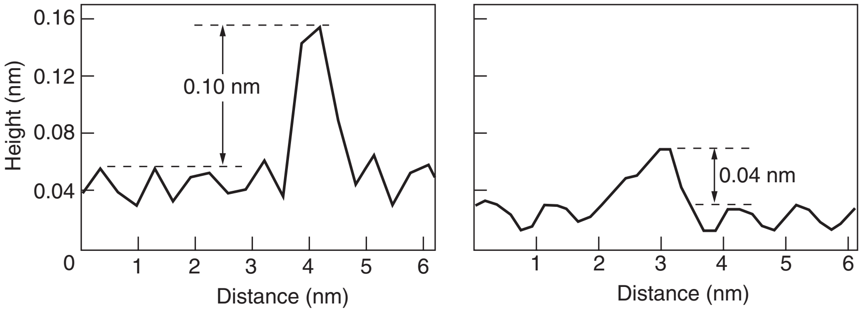


Figure 9. Incorporation of Phosphorus into the Surface

(a) This schematic diagram illustrates how phosphine molecules adsorb onto the bare Si(100)-(2 × 1) surface. The filled-state STM image in (b) is of a pair of adsorbed phosphine molecules, and (c) shows the line profile through the left molecule. (d) After annealing the surface to 400°C, the phosphorus atom incorporates into the silicon surface and forms a Si-P heterodimer. (e)–(f) These figures show the filled-state STM image of the heterodimer and the corresponding line profile. A comparison between (c) and (f) shows that there is a characteristic height difference between the nonincorporated and incorporated phosphorus, the former extending higher above the surface plane.

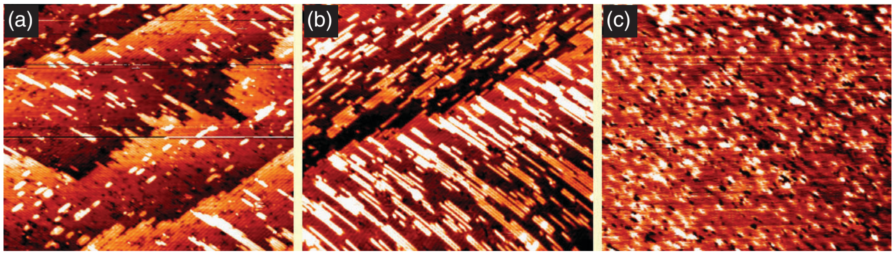


Figure 10. Images of Different Silicon Coverages

These images (100 nm × 85 nm) of a Si(100)-(2 × 1) surface show different stages of epitaxial silicon growth. The silicon was deposited while the sample remained in the microscope and was held at about 250°C. (a) The sample is shown after a 0.08 monolayer was grown. Epitaxial growth is demonstrated by the elongated shape of the islands and their direction being perpendicular to the underlying dimer rows. (b)–(c) The sample is shown after a deposition of 0.5 monolayer and a complete monolayer, respectively. At the growth temperature noted above, the surface is rough. Defects and silicon vacancies dominate the topography.

will try to grow the silicon at an elevated temperature. Because the layers will grow epitaxially, we can eliminate the subsequent anneal, but the challenge will be to find a growth temperature that also minimizes the segregation and diffusion of the phosphorus atoms.

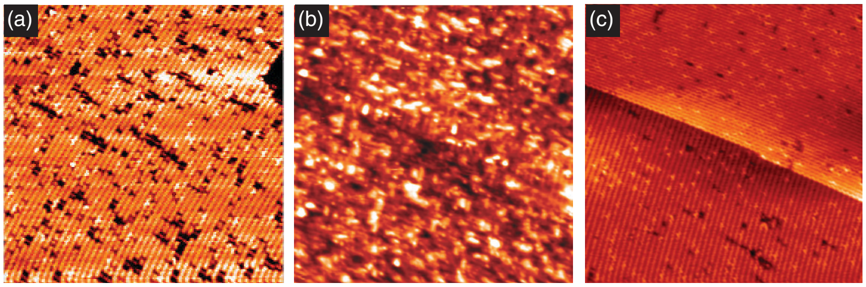


Figure 11. Silicon Overgrowth and Annealing after Low PH_3 Dosing
(a) A Si(100)-(2 × 1) surface is shown after low PH_3 dosing and annealing to incorporate phosphorus atoms into the Si-P heterodimers. The heterodimers are visible as bright zigzag structures. The image size is 50 × 50 nm^2. (b) The epitaxially overgrown surface is shown after annealing at 250°C. The image size is also 50 × 50 nm^2. The surface is too coarse for the SSQC and must be annealed. (c) After annealing at 600°C, the surface is flat. The bright spots indicate, however, that phosphorus has diffused to the surface. The image size is 55 × 55 nm^2.

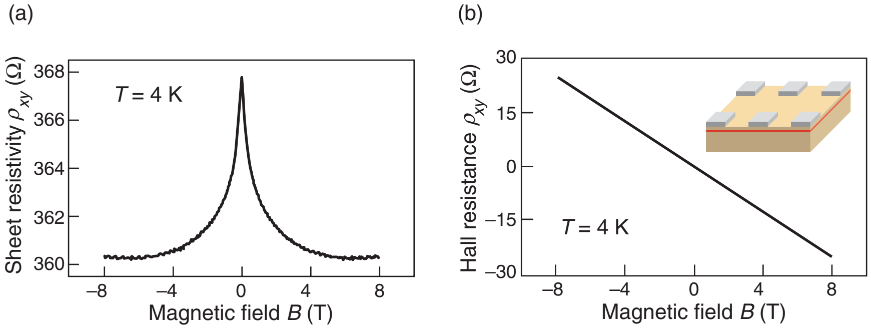


Figure 12. Electrical-Activity Tests
We wanted to check that phosphorus atoms incorporated in the silicon surface are electrically active. (a) The longitudinal resistivity ρ_{xx} of the delta-doped sample as a function of magnetic field was measured at 4 K. From this curve, a strong negative magnetoresistance is clearly observed, and it indicates the two-dimensional (2-D) nature of the delta-doped layer. (b) The Hall resistivity ρ_{xy} of the sample gives a 2-D carrier density of 2.0 × 10^{14} cm^{-2}. This number agrees with our dopant density and indicates that each phosphorus dopant is electrically active. The inset is a schematic of the phosphorus delta-doped silicon sample with metal surface contacts in the van der Pauw arrangement.

A significant number of experiments need to be conducted to determine the optimal encapsulation conditions. By integrating a small silicon evaporator into the STM chamber, we have already begun to study the epitaxial deposition of thin silicon layers at low temperatures. Figure 10 shows growth in the thickness of silicon of up to one monolayer at 250°C. The new layer grows epitaxially. Before it is annealed, the complete monolayer still exhibits vacancies that are not filled during the silicon overgrowth. Their possible detrimental effects on the operation of the quantum computer will have to be evaluated.

We have also begun to explore the first growth strategy (see Figure 11). We incorporated phosphorus into the silicon surface, deposited a few monolayers of silicon at room temperature, then annealed the sample for 1 minute at 250°C. As seen in Figure 11(b), this surface was fairly coarse and not suitable for subsequent epitaxial growth. A flat surface structure with island-free terraces was observed only after the sample had been annealed at 600°C. Figure 11(c), however, shows that, at those elevated temperatures, the phosphorus atoms have diffused to the surface. Although that result is disappointing, we are not discouraged. Ours are the first such studies of phosphorus encapsulation and silicon overgrowth. The preliminary results simply demand that we look for a new way to obtain a flat surface at lower annealing temperatures or an alternative way to inhibit phosphorus diffusion.

We have, however, settled the question of whether the incorporated phosphorus atoms are electrically active, that is, whether their donor electrons are free to conduct. We first grew a thin layer of phosphorus on a silicon substrate and buried it under a thick silicon layer (grown at the relatively low growth temperature of 250°C), creating a so-called delta-doped layer. According to the literature, our growth and annealing conditions resulted in a two-dimensional (2-D) density of 1.7 × 10^{14} phosphorus atoms per centimeter squared. If each atom is electrically active, it would contribute one free electron to the substrate. When we measured the electron density through the Hall effect at a sample temperature of 4 kelvins, the result was a 2-D density of 2.0 × 10^{14} electrons per centimeter squared (see Figure 12). As the two numbers agree within measurement errors, it seems that all the phosphorus atoms are electrically active (Oberbeck et al. 2002). This result suggests that the phosphorus atoms are incorporated in substitutional, rather than interstitial, sites, which is the ideal environment for the SSQC qubits.

Qubits, Defects, and Dopants

Although we have a clear strategy for creating and burying the phosphorus array, the volume surrounding the array in a working quantum computer must also be free from crystal impurities and defects. In general, defects disrupt the crystal structure and can create new pathways for quantum decoherence, which would inhibit qubit operations. Charged defects can be particularly disruptive. If the charge arises from an unpaired electron, then by necessity, there is an "impurity" spin that can interact with a qubit and affect its quantum state. Furthermore, the Coulomb potential of a charged defect lying close to a qubit can interfere with gate operations because it can offset the voltage applied to the qubit-controlling gate electrode.

Fortunately, the STM allows us to check the status of the buried qubits and charged defects during the fabrication of the quantum computer. Scanning tunneling microscopy is routinely used in characterizing the charge of individual defects found on the cleaved surfaces of compound semiconductors (Zheng et al. 1994, Lengel et al. 1994, Ebert et al. 1996). The charge becomes visible because of the so-called charge-induced band bending, illustrated in Figure 13. The states made available by band bending attract charge carriers that screen, or shield, the charged defect. Because bending shifts electronic states into or out of the window defining the source of the tunneling current, it produces a measurable enhancement or depression around the defect in the STM images. The characteristic length scale of this screening effect is given by the Debye screening length, which depends on the semiconductor's intrinsic properties: its dopant type and concentration (Dingle 1955).

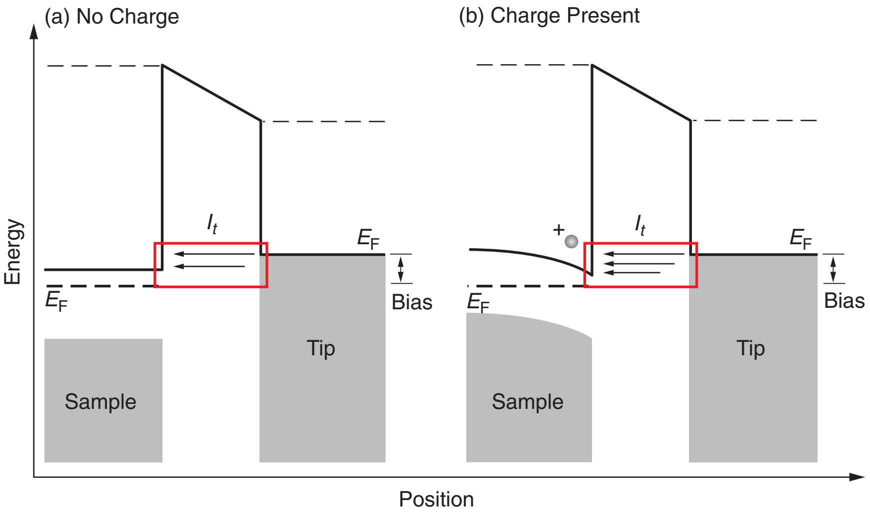


Figure 13. Band Bending

The electronic energy of an empty-state tunneling current is shown for a semiconducting sample with no charge (left) and with positive charge (right). The effects of surface states and defects have been neglected for illustration purposes. The charge-induced band bending shifts more states into the window between the Fermi levels of the tip and the sample. In this case, the increased state density relative to the rest of the neutral surface creates a long-range enhancement centered on the charge that falls off approximately like a screened Coulomb potential with a length scale set by the Debye screening length.

Table I. Expected Effect of Local Charge on Surface LDOS

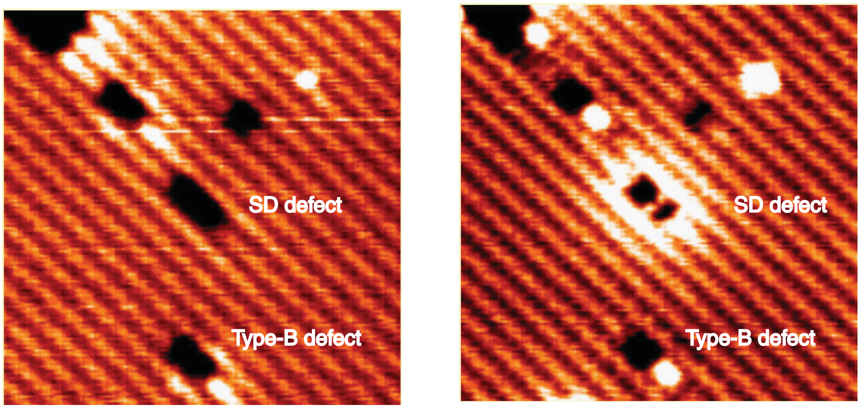
Silicon(100)-2 × 1 Surface	Imaging Condition	+Charge	–Charge
***n*-type**	Empty states Filled states	Enhanced No effect	Depressed Enhanced
***p*-type**	Empty states Filled states	Enhanced Depressed	No effect Enhanced

These techniques for imaging charge have not been demonstrated on silicon surfaces until now because it has been generally assumed (based on techniques such as photoelectron spectroscopy that probe large surface areas) that the Fermi level at the surface of silicon is pinned. If that assumption is true, the bands cannot respond to charge near the surface. But by taking into account what occurs locally and by drawing on other results obtained with the STM, we have determined that pinning of the Fermi level does not occur for clean Si(100)-(2 × 1) surfaces, except in the vicinity of type-C defects. This has allowed us to image charged defects on these clean surfaces for the first time (Brown et al. 2002).

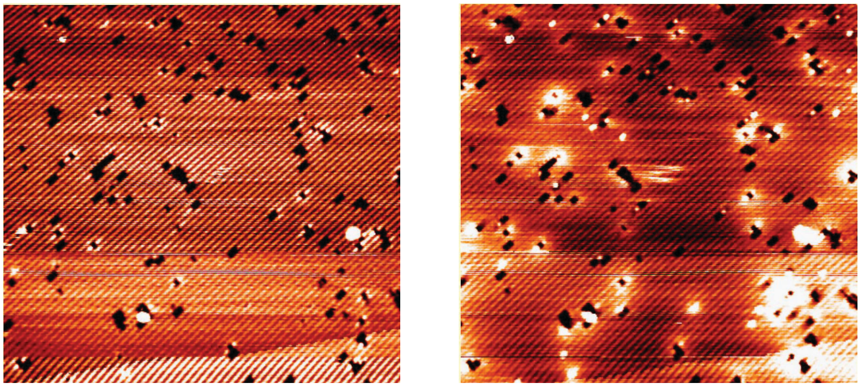
Considering the band structure as it is currently understood, we can qualitatively determine which types of charge should be detectable in filled- and empty-state imaging on a clean silicon(100)-(2 × 1) surface for both *p*-and *n*-type materials.[3] These predictions, made under the assumptions of nondegenerate doping, a tip work function of 3 to 4 electron volts, and a low C-defect density, are compiled in Table I. As noted in the table, under some conditions, we anticipate no change in the appearance of an STM image. That result is singularly different from what is seen on compound semiconductors and arises from surface states derived from the π-bond. These states, which are not present on the compound semiconductor surfaces, limit the amount of band bending that can occur.

Based on the expectations listed in Table I, we performed STM experiments at sample biases between ±1.5 volts on clean (2 × 1) surfaces of Si(100) samples doped with phosphorus (approximately 8 × 10^{15} phosphorus atoms per cubic centimeter). Low sample biases were used to ensure that effects near the band edges (for example, band bending) contributed strongly to the tunneling current. In these experiments, we were able to image charged defects consistent with our *n*-type predictions.

One such charged defect is shown in Figure 14. This defect is commonly observed in studies of Si(100)-(2 × 1) surfaces on thermally prepared samples and is typically referred to as a split-off dimer (SD, also called the 1+2 DV) defect. It consists of an A- and a B-defect on the same row, separated by one intact dimer. The empty-state image of the SD defect shows a long-range perturbation, but the filled-state image shows no corresponding feature even though the filled-state imaging is closer to the valence band edge on this *n*-type material.


(a) Filled-State Image

(b) Empty-State Image


(c) Filled-State Image (low resolution)

(d) Empty-State Image (low resolution)

Figure 14. Finding Charged Defects

(a) This filled-state image shows a charged split-off dimer (SD) defect on a Si(100)-(2 × 1) surface (center) and a type B defect (bottom center). (b) The same defects are shown for an empty-state image. The bright "cloud" in the empty states around the central defect indicates that there is an associated positive charge. The other defects in the image appear neutral. These images measure 16.8 nm × 16.8 nm and were acquired "simultaneously" by raster scans interleaved at each bias. The asymmetric nature of the screening signature in the two biases is consistent with our expectations. (c)–(d) Pictured here are filled- and empty-state images from a different sample, taken at lower resolution. Only about one-third of the defects are charged. The images are 57 nm × 55 nm.

These results are consistent with expectations based on Table I, indicating positive charge associated with the SD defect. The enhanced density of states—bright region in Figure 14(b)—appears to be nearly radially symmetric and approximately centered on the defect structure. Sections through the data show that the signature is discernible out to about 4.5 nanometers from the center. The corresponding Debye screening length, obtained from a screened Coulomb potential function fit to the sections, is approximately 3 nanometers. That result was unexpected. The typical bulk value for the screening length that is consistent with our dopant density (which correlates with the number of charge carriers) is several tens of nanometers. The short screening length indicates a high dopant density at the surface. One explanation is that, on our thermally prepared surface, buried dopants may diffuse because of the high temperature, and the actual density at the surface could be high enough to account for the 3-nanometer screening length.

---

[3] Electrical conduction in *n*-type materials is associated with electrons. In *p*-type materials, it is associated with holes.

Another interesting result is that, for samples with defect densities less than 5 percent, only about one-third of the surface defects are charged. This finding tells us that the charge is not associated with the simple vacancy structure observed in the images but must arise from more subtle effects. Charged defects may be due to, for example, rebonding differences among second-layer atoms. Charged and neutral defects may also coexist because of subsurface impurities or gas-phase species adsorbed in the vacancy structure itself. At this point and by using only scanning tunneling microscopy, we are unable to ascertain why only some defects are charged.

The fact that charge can be imaged on a silicon surface tells us that, after creating a flat overlayer, we will be able to detect the subsurface charged qubits. This finding is important for determining whether the qubits move during subsequent silicon growth. And looking beyond the Kane architecture, our results will be applicable to any implementation of a solid-state, silicon-based quantum computer.

Future Challenges

To date, we have demonstrated most of the individual steps required to successfully fabricate the Kane SSQC. We can create a small phosphorus array (O'Brien et al. 2001) and incorporate that array into the silicon surface. We have shown that the phosphorus atoms remain electrically active (oberbeck et al. 2002). We can grow silicon epitaxially in the STM at a temperature that should leave the array intact, and we can detect charged defects at the surface. Although not reported in this article, the Semiconductor Nanofabrication Facility housed at the University of New South Wales in Sydney, Australia, has fabricated metallic gates with dimensions close to those required for proper operation of the quantum computer.

As we integrate the aforementioned steps and try to produce a few-qubit device, several questions remain to be answered. Will the qubit array stay intact during silicon overgrowth and during any required postanneals? Can we remove defects during fabrication and, if not, to what extent will vacancies or impurities affect the computer operation? Will we introduce charge defects at the interface between the silicon overlayer and the insulating layer? How well can we register the gates with the qubits, once the array has been built?

Still, the number of questions that confront us today is far smaller than the number that faced us three years ago, when we first contemplated the steps involved in fabricating the SSQC. At that time, each question was tied to a long list of experimental obstacles that needed to be overcome. Through the combined efforts of two laboratories in the United States and Australia, we have been able to develop experimental procedures that have moved us closer to fabricating a qubit array. Given our prior success, we are hopeful that the remaining issues can be addressed successfully as well.

On a different note, one exciting idea that has emerged recently is the possibility that STM can detect single spins. Yshay Manassen et al. (2000) reported detection of a spin-induced alternating-current component in the STM tunneling current. Recent theoretical work, discussed in the article "Theory of Single-Spin Detection with a Scanning Tunneling Microscope" on page 184, offers an explanation and puts the experimental finding on firmer ground. At Los Alamos, we are in the process of modifying the electronics of our STM and adding an external magnetic field with the hope of confirming the effect. If we are successful, directly studying spin-spin interactions and creating, manipulating, and reading out surface-bound qubits may become reality. Such a possibility is indeed exciting. ■

Acknowledgments

This work would not have been possible without the effort of many students and postdoctoral researchers. They include Jeremy O'Brien, Steven Schofield, Neil Curson, Lars Oberbeck, and Holger Grube.

Further Reading

Bardeen, J. 1961. Tunneling from a Many-Particle Point of View. *Phys. Rev. Lett.* **6**: 57.

Brown, G. W., H. Grube, M. E. Hawley, S. R. Schofield, N. J. Curson, M. Y. Simmons, and R. G. Clark. 2002. Imaging Charged Defects on Clean Si(100)-(2 ×1) with Scanning Tunneling Microscopy. *J. Appl. Phys.* **92**: 820.

Chander, M., Y. Z. Li, J. C. Patrin, and J. H. Weaver. 1993. Si(100)-(2 × 1) Surface Defects and Dissociative and Nondissociative Adsorption of H_2O Studied with Scanning Tunneling Microscopy. *Phys. Rev. B* **48**: 2493.

Chen, C. J. 1990. Origin of Atomic Resolution on Metal Surfaces in Scanning Tunneling Microscopy. *Phys. Rev. Lett.* **65**: 448.

Dingle, R. B., 1955. *Phil. Mag.* **46**: 831.

Ebert, P., X. Chen, M. Heinrich, M. Simon, K. Urban, and M. G. Lagally. 1996. Direct Determination of the Interaction between Vacancies on InP(110) Surfaces. *Phys. Rev. Lett.* **76**: 2089.

Ebert, P., M. Heinrich, M. Simon, C. Domke, K. Urban, C. K. Shih et al. 1996. Thermal Formation of Zn-Dopant-Vacancy Defect Complexes on InP(110) Surfaces. *Phys. Rev.* B **53**: 4580.

Kane, B. E. 1998. A Silicon-Based Nuclear Spin Quantum Computer. *Nature* **393**: 133.

Lengel, G., R. Wilkins, G. Brown, M. Weimer, J. Gryko, and R. E. Allen. 1994. Geometry and Electronic Structure of the Arsenic Vacancy on GaAs(110). *Phys. Rev. Lett.* **72**: 836.

Liu, L., J. Yu, and J. W. Lyding. 2001. Atom-Resolved Three-Dimensional Mapping of Boron Dopants in Si(100) by Scanning Tunneling Microscopy. *Appl. Phys. Lett.* **78**: 386.

Lyding, K. W., T.-C. Shen, J. S. Hubacek, J. R. Tucker, and G. C. Abeln. 1994. Nanoscale Patterning and Oxidation of H-Passivated Si(100)-(2 × 1) Surfaces with an Ultrahigh Vacuum Scanning Tunneling Microscope. *Appl. Phys. Lett.* **64**: 2010.

Manassen, Y., I. Mukhopadhyay, and N. R. Rao. 2000. Electron-Spin-Resonance STM on Iron Atoms in Silicon. *Phys. Rev. B* **61**: 16223.

McEllistrem, M., G. Haase, D. Chen, and R. J. Hamers. 1993. Electrostatic Sample-Tip Interactions in the Scanning Tunneling Microscope. *Phys. Rev. Lett*. **70**: 2471.

Oberbeck, L., N. J Curson, M. Y. Simmons, R. Brenner, A. R. Hamilton, S. R. Schofield, and R. G. Clark. 2002. Encapsulation of Phosphorus Dopants in Silicon for the Fabrication of a Quantum Computer. (To be published in *Appl. Phys. Lett.)*

O'Brien, J. L., S. R. Schofield, M. Y. Simmons, R. G. Clark, A. S. Dzurak, N. J. Curson et al. 2001. Towards the Fabrication of Phosphorus Qubits for a Silicon Quantum Computer. *Phys. Rev.* B **64**: 161401(R).

Owen, J. H. G., D. R. Bowler, C. M. Goringe, K. Miki, and G. A. D. Briggs. 1995. Identification of the Si(001) Missing Dimer Defect Structure by Low Bias Voltage STM and LDA Modeling. *Surf. Sci.* **341**: l1042.

Selloni, A., P. Carnevali, E. Tosatti, C. D. Chen. 1985. Voltage-Dependent Scanning-Tunneling Microscopy of a Crystal Surface: Graphite. *Phys. Rev.* B 31: 2602.

Tersoff, J., and D. R. Hamann. 1985. Theory of the Scanning Tunneling Microscope. *Phys. Rev.* B **31**: 805.

Ukraintsev, V. A., Z. Dohnalek, and J. T. Yates, Jr. 1997. Electronic Characterization of Defect Sites on Si(001)-(2 × 1) by STM. *Surf. Sci.* **388**: 132.

Yates, J. T. Jr., and V. A. Ukraintsev. 1996. The Role of Nickel in Si(001) Roughening. *Surf. Sci.* **346**: 31.

Zheng, J. F., X. Liu, N. Newman, E. R. Weber, D. F. Olgetree, and M. Salmeron. 1994. Scanning Tunneling Microscopy Studies of Si Donors (Si_{Ga}) in GaAs. *Phys. Rev*. Lett. **72**: 1490.

Marilyn E. Hawley received her Ph.D. in physics from The Johns Hopkins University in 1987. Marilyn joined Los Alamos National Laboratory as a postdoctoral fellow in 1989. She was the first to successfully image and identify the spiral growth mechanism in high-temperature superconducting films using scanning tunneling microscopy (STM), results that were featured on one of the covers of Science. Marilyn has over 18 years of experience in STM and many advanced atomic force microscopy techniques. In 1991, she became a staff member in the Center for Materials Science at Los Alamos. Later, Marilyn established the Scanning Probe Microscopy Laboratory, a facility devoted to the development and use of scanning probe techniques for research on a broad spectrum of materials. She currently leads the Los Alamos STM team in an effort aimed at fabricating a qubit array for a possible future solid-state quantum computer.

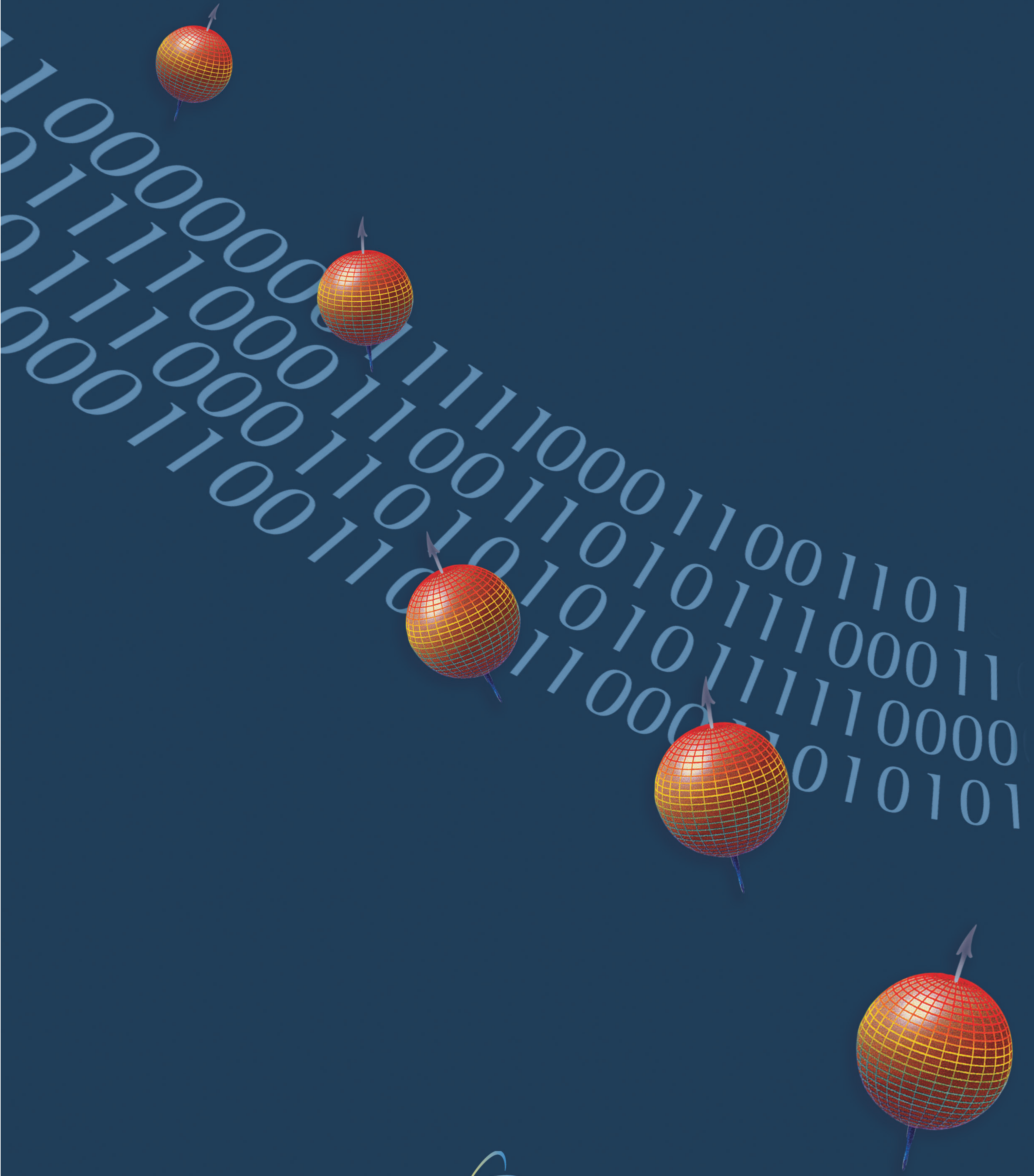
Geoff Brown received a bachelor's degree in physics from Abilene Christian University and a Ph.D. in physics from Texas A&M University, where he studied the physics and chemistry of semiconductor surfaces using scanning tunneling microscopy (STM). In 1996, Geoff started at Los Alamos as a postdoctoral fellow in the Center for Materials Science, and he is now a technical staff member in the Scanning Probe Microscopy Laboratory at Los Alamos. His research involves STM and atomic-force microscopy studies of a wide range of materials, including semiconductors, complex oxides, metals, and superconductors.

Michelle Y. Simmons is currently the Director of the Atomic Fabrication Facility and a Queen Elizabeth II Research Fellow at the University of New South Wales, in Sydney, Australia. She joined the Centre for Quantum Computer Technology as the Program Manager in Atomic Fabrication and Crystal Growth in 2000, after completing a postdoctoral fellowship at the University of Cambridge, in the United Kingdom, where she was in charge of the design, fabrication, and characterization of ultrahigh-quality quantum electronic devices. She has over 14 years of experience in all aspects of semiconductor crystal growth, device fabrication, and electrical characterization of quantum electronic devices. Her current research interests are to understand how quantum electronic devices work as they become purer and smaller and to use this knowledge toward building the next generation of devices by using quantum principles—in particular a silicon-based quantum computer.

For the biography of Robert Clark, see page 300.

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the US Department of Energy under contract W-7405-ENG-36. All company names, logos, and products mentioned herein are trademarks of their respective companies. Reference to any specific company or product is not to be construed as an endorsement of said company or product by the Regents of the University of California, the United States Government, the US Department of Energy, nor any of their employees. The Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

LA-UR-02-4969



Los Alamos
NATIONAL LABORATORY